PUBLICATIONS DE L'INSTITUT DE STATISTIQUE
DE L'UNIVERSITÉ DE PARIS

MÉMOIRES ET CONFÉRENCES SUR LE CALCUL DES PROBABILITÉS,
LA STATISTIQUE THÉORIQUE ET APPLIQUÉE, L'ÉCONOMÉTRIE

STATISTICAL METHODS AND SCIENTIFIC INFERENCE
SIR R. A. FISHER
Compte-Rendu présenté par J. R. Barra

G. BODIOU
PROBABILITÉ SUR UN TREILLIS NON MODULAIRE

P. THIONET
SUR LES AGRÉGATS DE DISTRIBUTIONS STATISTIQUES D'UNE MÊME FAMILLE

CARLO BONFERRONI
LES VALEURS MÉDIANES ET LA THÉORIE DE LA MESURE

ANALYSES D'ARTICLES

VOL. VI - FASCICULE 1 - 1957

PARIS
11, Rue Pierre Curie

Toute la correspondance relative aux publications
doit être envoyée à l'adresse :

INSTITUT DE STATISTIQUE DE L'UNIVERSITE DE PARIS

INSTITUT HENRI POINCARÉ - 11, Rue Pierre Curie - PARIS (V^e)

Les manuscrits doivent être envoyés à *M. Daniel DUGUÉ*,
à l'adresse précédente.

Abonnements :	Pour la France 1.700 francs français Pour l'Etranger 2.000 francs
Vente au numéro :	*(fascicule de 50 pages environ)* Pour la France 500 francs Pour l'Etranger 600 francs

UBLICATIONS DE L'INSTITUT DE STATISTIQUE
DE L'UNIVERSITÉ DE PARIS

MÉMOIRES ET CONFÉRENCES SUR LE CALCUL DES PROBABILITÉS,
LA STATISTIQUE THÉORIQUE ET APPLIQUÉE, L'ÉCONOMÉTRIE

STATISTICAL METHODS AND SCIENTIFIC INFERENCE

SIR R. A. FISHER
Compte-Rendu présenté par J. R. Barra

G. BODIOU

PROBABILITÉ SUR UN TREILLIS NON MODULAIRE

P. THIONET

SUR LES AGRÉGATS DE DISTRIBUTIONS STATISTIQUES D'UNE MÊME FAMILLE

CARLO BONFERRONI

LES VALEURS MÉDIANES ET LA THÉORIE DE LA MESURE

ANALYSES D'ARTICLES

VOL. VI - FASCICULE 1 - 1957

PARIS
11, Rue Pierre Curie

"STATISTICAL METHODS AND SCIENTIFIC INFERENCE"

SIR R. A. FISCHER

Compte-Rendu présenté par J. R. Barra

Examiner les bases logiques de l'analyse statistique et ses modes de raisonnement, tel est le but que se propose Sir R. Fisher dans son dernier livre : "Statistical Methods and Scientific Inference".

Et certes le besoin s'en faisait sentir; si dans l'immense développement des statistiques scientifiques ou appliquées l'auteur note deux époques : d'abord essentiellement avec Pearson et Galton l'étude mathématique des problèmes, puis plus récemment les essais d'expériences planifiées, il n'en affirme pas moins que la construction d'expériences probantes et leur interprétation doivent s'intégrer dans un ensemble logique et cohérent : notre mode d'investigation du monde réel. Il s'agit d'établir les bases même du raisonnement inductif statistique, n'hésitons donc pas à assurer parfaitement nos étapes de raisonnement.

Le plan suivi par Sir R. Fisher est le suivant :

A - LA TENTATIVE DE BAYES des probabilités à priori et les critiques qui lui ont été faites. Pour l'auteur, ces critiques montrent que la même forme de raisonnement ne pourra pas être appliquée dans toutes les situations et il conviendra, suivant les cas, d'utiliser, soit la méthode de Bayes, soit un test de signification, soit ...

B - QUELQUES FORMES GENERALES DE RAISONNEMENT :

Le test d'une hypothèse et la valeur qu'il convient de lui donner, puis la méthode fiduciaire d'estimation.

C - L'étude comparée de la notion de probabilité et du concept de vraisemblance sur des exemples.

D - Problèmes particuliers à l'estimation.

Au cours de son ouvrage, l'auteur fait de nombreuses incidences sur la notion de probabilité, tant pour remarquer la diversité de significations pratiques que chaque esprit lui donne, que pour justifier la méthode des probabilités fiduciaires. Nous analyserons ces remarques dès maintenant, pour apporter quelques éclaircissements à la suite.

I. - LA NOTION DE PROBABILITÉ

Les points essentiels pour Sir R. Fisher semblent être les suivants :

α - La définition mathématique n'est qu'un ensemble d'axiomes à utiliser quand nécessaire (p. 32) la théorie axiomatique n'a jamais été prise bien au sérieux dans les branches appliquées (p. 107). Le facteur d'incertitude y est inopérant (p. 109).

β - Pour appliquer les résultats mathématiques on doit prendre certaines précautions : par exemple, pour qu'un joueur en expectative devant le jet d'un dé, soit obligé d'admettre la probabilité 1/6, il faut et il suffit qu'il n'ait aucune possibilité pour distinguer entre elles les épreuves futures, donc que son ignorance complète soit bien établie (p. 32).

γ - En remarquant le rôle de l'ignorance comme donnée, l'auteur est amené à critiquer la conception de Laplace de la probabilité. Renouvellant contre lui l'objection de confusion entre possible et probable - et que d'ailleurs tous les auteurs ne trouvent pas aussi importante - Sir R. Fisher estime que la définition Laplacienne de la probabilité est exprimée de telle façon que l'on puisse confondre les deux propositions :

a) J'attriburai la même probabilité à un évènement au cours d'épreuves futures, car je ne peux rien retenir pour les distinguer.

b) Je considérerai des hypothèses comme également probables si je ne peux rien retenir pour les distinguer (p. 34).

Quand l'auteur pense à la théorie axiomatique, il semble que ce soit à la théorie des fonctions complètement additives d'ensemble, cependant on ne saurait oublier qu'avant de considérer un évènement comme élément d'un ensemble muni d'une loi de probabilité, on a dû, non seulement faire quelques abus de langage du type : confondre probabilité pour que la proposition "à la suite de telle épreuve l'évènement se produira" soit vraie, avec probabilité de l'évènement par rapport à l'épreuve considérée, mais surtout préciser la notion de catégorie d'épreuves, notion essentielle (Fréchet ...) qui doit être mise en évidence dans tous les cas délicats et que l'on peut considérer comme représentative du "facteur d'incertitude".

Quant à la condition (β) n'est-elle pas aussi nécessaire à la validité de la probabilité mathématique elle-même ? N'est-ce pas elle qui justifie l'assertion suivante : Si p est la probabilité de E sur la catégorie d'épreuves représentée par 1 seul jet, p^n est la probabilité de $E \times \overset{n}{\cdots} \times E$ sur la catégorie d'épreuves représentée par n jets ? Autrement dit l'indépendance des épreuves. Il apparaitrait donc que les propositions (a) et (b) de (γ) ne soient pas comparables. (a) justifie l'indépendance des épreuves (b) est relatif à un mode de formation de la loi de probabilité. Et certes en ce qui concerne (b), comme le montre Sir R. Fisher, les critiques de Venn et Boole, doivent servir à éviter un usage abusif de l'ignorance.

Ainsi les points (α) (β) (γ) prennent-ils toute leur valeur comme précautions à prendre en calcul des probabilités mais ne pourraient infirmer l'image mathématique que l'on peut actuellement construire des problèmes probabilistes ! Mais si l'on se préoccupe de rigueur, ne doit-on pas s'astreindre à ne parler de probabilité uniquement pour des éléments aléatoires ? Cette idée avait déjà conduit M. Fréchet à proposer probabilités des hypothèses au lieu de probabilités des causes. On pourra constater dans la suite qu'un tel souci évite bien des embarras.

II. - LE CONCEPT DE VRAISEMBLANCE ET LE PROBLÈME STATISTIQUE

Parallèlement à la notion de probabilité, Sir R. Fisher souligne avec une particulière vigueur l'importance de la notion de vraisemblance et même le rôle complémentaire de celle-ci par rapport à la probabilité dans la représentation mathématique de l'incertitude. Sir R. Fisher avait déjà remarquablement exprimé ceci dans "Inverse probability" dans les termes :

"Connaissant la population nous pouvons exprimer notre connaissance incomplète de l'échantillon à venir en termes de probabilité.

Connaissant l'échantillon nous pouvons exprimer notre connaissance incomplète de la population en termes de vraisemblance".

L'auteur note la différence entre ces deux concepts qui n'obéissent pas aux mêmes lois mais qui pourtant mesurent tous deux une certaine "croyance rationnelle". Déduire de là une schématisation mathématique simple du problème statistique il n'y a qu'un pas, nous le ferons brièvement ne serait-ce que pour préciser et simplifier le langage ; admettons la définition suivante :

> Soit un espace fonctionnel $\mathcal{H}$ de fonctions de probabilité que nous nommerons Hypothèses f ($f \in \mathcal{H}$). Il s'agit de tirer des conclusions sur $\mathcal{H}$, étant donnée une "fonction de vraisemblance sur $\mathcal{H}$", celle-ci définie comme suit
>
> A tout $f \in \mathcal{H}$ correspond une certaine population, sur toutes ces populations est définie une même catégorie d'épreuves. Supposons que sans que l'on sache sur quelle population on opère on ait appliqué l'épreuve et soit $\mathcal{E}$ le résultat. On appelera vraisemblance de f, $L_{\mathcal{E}}(f)$: "probabilité pour que l'épreuve étant appliquée à la population correspondant à f le résultat en soit $\mathcal{E}$".

On utilise quelquefois des quantités proportionnelles (vraisemblance relative) ou encore le logarithme de L.

Sir R. Fisher souligne souvent au cours de son ouvrage que la notion de probabilité, ne sera pas toujours susceptible de traduire mathématiquement l'incertitude d'une conclusion. En effet dans la définition précédente, les informations que l'on a sur $\mathcal{H}$ sont incomplètes et pourtant comment pourrait-on parler de loi de probabilité sur $\mathcal{H}$ puisque les éléments f ne sont pas aléatoires ? En effet possède-t-on une catégorie d'épreuves sur $\mathcal{H}$?

Notons qu'en pratique :

1° - Les données permettent de réduire $\mathcal{H}$, par exemple à un espace fonctionnel isomorphe à R (cas paramétrique).

2° - L'épreuve utilisée souvent est un échantillon, mais on utilise ensuite presque toujours une fonction T (statistique) de cet échantillon et il conviendra, par exemple, de savoir si T traduit bien les données expérimentales (quantité d'information).

3° - Suivant le but poursuivi on tirera des conclusions orientées dans un certain sens :

- test de signification, c'est-à-dire, choisissant un f particulier à $\mathcal{H}$ on étudie la discordance entre l'observation et ce qui aurait été observé si l'épreuve avait été appliquée à la population de f.
- estimation, choix d'un θ optimum, dans quel sens ? etc ...

- processus de décision .

Sir R. Fisher examine séparément ces diverses méthodes, sur un grand nombre d'exemples mais dégage particulièrement l'esprit du

III. - TEST DE SIGNIFICATION

Dans l'analyse que Sir R. Fisher fait de la structure d'un test de signification deux points lui paraissent essentiels.

A - Une vraisemblance très petite jouit de la même valeur psychologique qu'une probabilité très petite.

Supposons que $L(f_o)$ - f_o hypothèse à rejeter - soit très faible on est placé devant le dilemne suivant (p. 39),

- ou bien un hasard extraordinaire a joué,
- ou bien $\mathcal{E}$ n'est pas issu de la population correspondant à f_o.

Sir R. Fisher estime qu'on éprouve la même répugnance à accepter f_o dans ces conditions qu'à escompter un évènement de probabilité égale à $L(f_o)$ et déclare (p. 43) :

"Bien qu'identifiable à une condition psychologique de rejet de l'hypothèse, le sentiment introduit par un test de signification a pour base objective un état de probabilité ... et si $L(f_o)$ représente une mesure des fondements rationnels de refus qu'il engendre il est pourtant évident qu'aucune probabilité n'est établie sur l'hypothèse elle-même".

Ainsi on pourrait étendre aux vraisemblances très voisines de o, le sens pratique que Borel donnait aux probabilités très voisines de o.

B - Ne pas confondre, comme le font Neymann et Pearson, test de signification et processus de décision, et ceci pour les raisons suivantes :

1° - $L(f_o)$ petit entraine la quasi-certitude de se tromper avec une fréquence bien inférieure à $L(f_o)$. En effet si $\mathcal{E}$ est issu de f_o on se trompe avec la fréquence $L(f_o)$ et si $\mathcal{E}$ n'est pas issu de f_o on ne se trompe pas. Comme un expérimentateur ne choisit pas f_o d'une façon quelconque on voit qu'en pratique la fréquence d'évènements est très largement inférieure à $L(f_o)$. D'autre part pour un processus de décision, il faudrait connaitre les deux erreurs ce qui est inutile dans un test de signification puisque si l'on rejette f_o on se préoccupe peu de la probabilité de l'accepter à tort. De plus ces erreurs sont subordonnées à l'espace $\mathcal{X}$ choisi, or celui-ci est le fruit de l'imagination du statisticien et n'est pas comparable à la population bien déterminée et ayant un sens pratique d'un processus classique d'acceptation. Enfin quand le test s'applique à un ensemble d'hypothèses, les calculs précédents ne sont strictement plus possibles. Nous noterons cependant qu'un effort a été fait pour remédier à ceci avec les structures semblables.

2° - C'est enfin méconnaitre le caractère d'un test de signification que d'imposer rigidement une fréquence d'erreur. A la suite d'un test, un expérimentateur ne prend qu'une décision provisoire, jusqu'à plus ample informé, et suivant les cas se montrera plus ou moins strict sur $L(f_o)$. Et Sir R. Fisher s'indigne à l'idée qu'un chercheur ne soit qu'un automate sans responsabilité ni initiative au service d'une immense organisation à lois imposées.

Et l'auteur conclut en se défendant d'introduire des "fonctions de prix dans les travaux scientifiques comme on doit le faire dans un processus de décision; car on cherche des méthodes également convaincantes à tous les esprits libres, indépendamment des intentions que l'on pourrait avoir quand à l'utilisation du savoir obtenu".

Le test de signification étant en quelque sorte l'analyse immédiate de la vraisemblance, le deuxième thème de l'ouvrage de Sir R. Fisher est la comparaison entre la méthode fiduciaire et celle des probabilités a priori (Bayes) qui fournissent l'une et l'autre, dans des cas simples, une sorte de fonction de probabilité sur $\mathcal{H}$. Est-ce justifié ?

IV. - MÉTHODE DE BAYES

Sir R. Fisher a déjà exposé cette méthode dans "Inverse probability" :

Supposons que la population dont est tirée l'observation ait été elle-même tirée au hasard d'une super-population de distribution connue a priori d'élément ϕ (f), la distribution a posteriori est :

$$\frac{\Psi(f)\, L_\varepsilon(f)}{\int_{(f)} L_\varepsilon(f)\ \phi(f)}$$

De plus en l'absence d'information on prendra ϕ uniforme.

S'attachant plus particulièrement à la validité de cette méthode, Sir R. Fisher à la suite de Venn et Boole, estime que la distribution a priori doit correspondre à une véritable donnée. Exemple (p. 18) :

Soient des souris noires de deux types BB et Bb satisfaisant à

(I) BB + grise ⟶ noire

(II) Bb + grise ⟶ 50% de grises et 50 % de noires

(III) Bb + Bb ⟶ 25% BB, 50% de Bb et 25% de grises.

Soit une souris noire d'une portée III, avec une grise elle donne 7 noires, est-elle BB ou Bb?

probabilités a priori BB:1/3 Bb:2/3

vraisemblance BB : 1 Bb:1/128

probabilités a posteriori

$$\begin{cases} BB = \dfrac{1 \times 1/3}{1 \times 1/3 + 2/3 \times 1/128} = \dfrac{64}{65} \\[2ex] Bb = \dfrac{2/3 + 1/128}{1 + 1/3 + 2/3 \times 1/128} = \dfrac{1}{65} \end{cases}$$

La donnée : "La souris appartient à une portée III" justifie ici les probabilités a priori de 1/3 et 2/3 mais dire qu'en l'absence d'information elle avait les probabilités a priori de 1/2 et 1/2 reviendrait à traiter le problème comme si elle était issue d'une portée donnant des BB et Bb à 50% et serait donc abusif.

Mais ne faudrait-il pas insister sur le fait que l'esprit accepte une loi de probabilité sur $\mathcal{H}$ (ici réduit à deux propositions parce que le fait d'être BB est évidemment aléatoire pour une souris noire issue de III, sur l'épreuve facile à concevoir : choix d'une souris noire parmi les noires d'une portée III?

Sir R. Fisher traite des exemples, en particulier prédiction sur de nouvelles épreuves (Règle de Succession) mais souligne qu'on ne doit pas oublier - même si le paramètre a disparu du résultat - que ces déductions supposent la validité des probabilités a priori.

V. - MÈTHODE FIDUCIAIRE

A la méthode précédente Sir R. Fisher oppose sa méthode fiduciaire (p. 51) qui n'est valable qu'en l'absence complète de distribution a priori mais qui exige :

1° - Des observations "continues", c'est-à-dire pratiquement suffisamment précises pour être considérées comme des valeurs observées d'une variable continue (p. 50).

2° - Il existe une statistique exhaustive, c'est-à-dire :

$$L(f) = L(f)$$

pour ε pour T

Supposons encore que T soit estimation de θ et que si T est tirée d'une population de paramètre θ on ait :

$$P = \text{prob}\ (T < t) = F(\theta, t)$$

Avec F continue et monotone par rapport à θ et t, et exposons avec Sir R. Fisher la méthode sur un exemple.

Soit une source radioactive, émettant des particules à des intervalles de temps x, indépendants, aléatoires, distribuées en $\theta e^{-\theta x} dx$, θ inconnu. Mesurons n intervalles x_i et posons :

$$T = \frac{n}{X} \qquad X = \sum_i x_i$$

on a

$$L(\theta) = \theta^n \prod_i e^{-\theta x_i} dx_i = \theta^n e^{-\theta X} \prod_i dx_i$$

prenons la vraisemblance relative

$$\ell(\theta) \theta^n e^{-\theta X} \text{ maximum pour } \theta = \frac{X}{n}$$

d'autre part la distribution de X est

$$\theta^n e^{-\theta X} \frac{X^{n-1}}{(n-1)!} dX \qquad (1)$$

d'où si l'on pose

$$\chi^2 = 2\,\theta X = 2n\frac{\theta}{T}$$

χ a une distribution d'après (1) qui est simple et même tabulée d'où :

quelque soit P, il existe $\chi^2(P)$ tel que

$$\text{Prob}\left[\chi^2 > \chi^2(P)\right] = P$$

soit

$$\text{Prob}\left[2n\frac{\theta}{T} > \chi^2(P)\right] = P$$

(Remarquons, en aparté, qu'ici χ^2 est aléatoire sur la catégorie d'épreuves complexe : mesure de n intervalles).

Sir R. Fisher en déduit alors la proposition

$$\text{Prob}\left[\theta > \frac{T}{2n}\chi^2(P)\right] = P$$

Ce qui représente à son avis une distribution sur θ. Comme justification nous nous limiterons aux propres termes de Sir R. Fisher (p. 54).

1° "Dans l'égalité (1) l'ensemble de référence est l'ensemble de toutes les valeurs de T pouvant être obtenues pour une valeur θ. Il, (1), a été démontré pour tout θ et est ainsi applicable à tous les couples (T, θ) obtenus pour toutes les valeurs de θ. Le couple particulier d'un expérimentateur appartient à cet ensemble élargi, à l'intérieur duquel la proportion des cas satisfaisant l'inégalité

$$\theta > \frac{T}{2n} X^2(p) \text{ est égale à P".}$$

2° - p. 56 "En conséquence il est essentiel d'introduire l'absence d'information a priori comme une donnée distincte pour démontrer complètement l'applicabilité de la méthode fiduciaire de raisonnement aux cas particuliers expérimentaux auxquels elle est appliquée".

Sir R. Fisher note plus loin qu'avec les notations du début $\frac{\partial F}{\partial T}dT$ est la vraisemblance de θ et $-\frac{\partial F}{\partial A}d\theta$ la "distribution fiduciaire" élémentaire de θ; et enfin précise qu'il ne faut pas confondre les résultats issus de la "distribution fiduciaire" avec les limites de tolérance.

Nous nous limiterons à indiquer les exemples largement traités par Sir R. Fisher en application des deux méthodes précédentes.

- étude d'un "échantillon normal" paramètres inconnus (p. 78) distribution de t et z,
- un cas de régression linéaire (p. 82),
- prédiction fiduciaire et Bayesanne (p. 111 à 121) et encore de nombreux autres.

Enfin dans son dernier chapitre Sir R. Fisher traite de la :

V. - THÉORIE DE L'ESTIMATION

Rappelant ses travaux précédents l'auteur précise quelques notions particulières à cette branche des statistiques : c'est-à-dire comment construire une statistique - définie à partir des observations - qui puisse être considérée comme une estimation satisfaisante. 3 notions essentielles.

A - CRITERE DE CONSISTANCE

Remarquons d'abord que si T estime θ, $f(T)$ doit estimer $f(\theta)$ d'où rejeter la condition de non-biaisage.

Ensuite si $T(a_i)$ est une estimation de $\varphi(A_i)$

a_i observations

θ_i paramètres

$T(\overline{a_i}) = \varphi(\theta_i)$ défini une estimation consistante.

Sir R. Fisher montre que cette définition est meilleure que la convergence en probabilité de $T \longrightarrow \varphi$ cas si :

T'_{N_1} est une statistique quelconque définie pour N_1 observations,

T'_N est une statistique asymptotiquement consistante, toujours définie.

$$\frac{1}{N}\left(N_1\, T'_{N_1} + (N-N_1)\, T_{N-N_1}\right)$$

satisfait à la condition asymptotique mais coincide avec T'_{N_1}, pour $N = N_1$.

B - CRITERE D'EFFICACITE

La première définition donnée fut asymptotique. Soit T une statistique dont la distribution devient normale si l'effectif n de l'échantillon tend vers l'infini, V(T) la variance de T.

Si nV(T) a une limite minimum T est efficace.

Un souci semblable pour des échantillons finis a conduit au critère de suffisance et à utiliser :

C - QUANTITE D'INFORMATION

T traduit-il bien toute l'information contenue dans l'échantillon en vue de $\varphi(\theta_i)$ à estimer? Cette question est résolue par la quantité d'information de Sir R. Fisher qui ici examine plus particulièrement le problème des statistiques (ou de l'information) subsidiaires et insiste encore sur l'importance de la vraisemblance dans tous ces problèmes.

PROBABILITÉ SUR UN TREILLIS NON MODULAIRE

(Calcul quantique des probabilités)

par

G. BODIOU

A. - BUTS ET RÉSULTATS DE CE MÉMOIRE

1°) Rappelons d'abord quelques caractères du formalisme quantique :

A tout vecteur unitaire, Ψ, d'un espace de Hilbert, H, correspond une pondération des multiplicités linéaires fermées ou m.l.f de H, telles que A, par :

$P(A/\Psi) = \| E(A)\Psi \|^2$, où E(A) est l'opérateur de projection orthogonale sur A.

On prend aussi en considération une pondération conditionnée par la m.l.f.B et définie par :

$$P(A/B,\Psi) = P(A/E(B)\Psi) = \| E(A).E(B).\Psi \|^2 / \| E(B)\Psi \|^2$$

Ces définitions ne dépendent que de la multiplicité d'ordre 1 qui supporte Ψ et tout le formalisme se définit en termes du treillis des multiplicités linéaires fermées de H, ou treillis (m.l.f).

Cette pondération quantique n'est une probabilité classique que sur des sous treillis particuliers de (m.l.f). En effet :

a) $A \cap B = \emptyset$ n'implique pas $P(A \cup B/\Psi) = P(A/\Psi) + P(B/\Psi)$, c'est-à-dire que le principe des probabilités totales n'est pas vérifié par la pondération quantique.

b) Comme $E(A).E(B) \neq E(A \cap B)$ (sauf si E(A) et E(B) commutent) le principe des probabilités composées n'est pas vérifié par les pondérations conditionnées quantiques. On interprète, d'un point de vue classique cette non-vérification en considérant que le "conditionnement quantique" est, en réalité, un "changement de catégorie d'épreuves". Il faut cependant constater que, s'il en est ainsi, le formalisme quantique établit, entre catégories distinctes d'épreuves, par sa pondération conditionnée, un lien ignoré des classiques.

c) On peut montrer que l'impossibilité de définir sur (m.l.f) une probabilité classique est impliquée par la structure de ce treillis.

Ce treillis est non-modulaire : c'est-à-dire que, sur lui, on peut avoir, à la fois :

$$A \leqslant C \quad \text{et} \quad A \cup (B \cap C) \neq (A \cup B) \cap C; \text{ (voir (1)}.$$

alors que, sur un treillis modulaire :

$$A \leqslant C \Longrightarrow A \cup (B \cap C) = (A \cup B) \cap C \text{ ; } (\Longrightarrow \text{marque l'implication})$$

Cette dernière relation exprime une distributivité restreinte, la distributivité générale s'exprimant par :

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Or, si un treillis supporte une probabilité classique, P(A), cette probabilité satisfait à la "relation modulaire" :

$$P(x) + P(y) = P(x \cup y) + P(x \cap y),$$

d'où l'on déduit aisément : $P(A \cup (B \cap C)) = P((A \cup B) \cap C)$, quand $A \leqslant C$;

et comme, terme à terme : $A \cup (B \cap C) \leqslant (A \cup B) \cap C$, on a :

$$P((A \cup B) \cap C - A \cup (B \cap C)) = 0,$$

pour toute loi de probabilité classique sur ce treillis.

Si l'on impose, alors, à ce treillis, d'être tel qu'il n'y ait que le minorant universel, $\emptyset$, dont la probabilité soit nulle pour toute loi, ce qui est vérifié par la pondération quantique sur (m.l.f) et par les probabilités classiques, il suit de la relation précédente que :

$$(A \cup B) \cap C - A \cup (B \cap C) = \emptyset,$$

c'est-à-dire que le treillis est nécessairement modulaire. (m.l.f), qui n'est pas modulaire, ne peut donc supporter une telle probabilité.

2°) Notre but est d'affaiblir convenablement les conditions imposées au calcul classique des probabilités pour que le support du nouveau calcul puisse être un treillis non-modulaire, tout en sauvegardant les notions probabilistes fondamentales : probabilités totales, probabilités conditionnelles, variables aléatoires, fonctions de répartition, indépendance, lois des grands nombres ...

Le treillis que nous décrivons est plus général que le treillis (m.l.f). Son immersion dans ce dernier est rendue possible par des axiomes supplémentaires dont le plus caractéristique est que tout élément du treillis est l'union d'un ensemble d'atomes dont la puissance est, au plus, infinie dénombrable. Mais ces restrictions sont artificielles et le formalisme probabiliste quantique peut être effectivement généralisé.

La trace du calcul élargi sur les sous-treillis booléens du treillis total est un calcul classique.

L'une des interprétations possibles du formalisme ainsi construit est la suivante :

L'ensemble des propositions expérimentales relatives à un système matériel donné est réparti en systèmes hypothético-déductifs : ce sont les divers treillis de certitudes. L'ensemble, supposé complet, des axiomes de l'un de ces systèmes étant symbolisé par son atome de base.

L'affirmation de vérité de l'un de ces systèmes se traduit par l'attribution de la probabilité 1 à son atome de base, et, par suite à tous ses éléments.

Une telle affirmation de vérité détermine la probabilité de toutes les propositions expérimentales.

Le conditionnement s'interprète alors comme suit : L'affirmation de vérité d'une proposition expérimentale n'appartenant pas au système déductif qui fixe la loi de probabilité, implique le remplacement de ce système par

celui, parmi ceux auxquels appartient la proposition considérée, dont la base axiomatique est la plus probable dans la pondération initiale.

Le calcul des probabilités généralisé serait une théorie de la pondération relative des systèmes hypothético-déductifs les uns par les autres.

Conclusion : Quel que soit l'aboutissement des travaux actuellement en cours pour atteindre une réalité plus profonde que celle que décrit la statistique quantique, cette statistique restera valable à son niveau propre.

Nous sommes convaincus qu'elle implique une généralisation fructueuse de la notion même de probabilité.

B. - PROBABILITÉS SUR UN TREILLIS NON-MODULAIRE

La consistance des axiomes est assurée par leur vérification sur (m.l.f). Leur indépendance n'est pas démontrée.

Les preuves des théorèmes et quelques vérifications d'axiomes sur (m.l.f) ont été reportées à la fin du mémoire, afin de mieux dégager les caractères essentiels du formalisme.

Les axiomes sont notés : A ..., les définitions : D ..., les théorèmes : Th ...; ceux d'entre eux dont la démonstration, ou quelque commentaire à leur sujet, figure en fin de mémoire, sont soulignés : A, D, Th.

I. TREILLIS ORTHOGONAL T.

T admet un élément nul, $\emptyset$, et un élément universel, u. Sur T est définie une complémentation, $A \longrightarrow A'$, dite orthocomplémentation parce qu'on lui impose la propriété : $A \leqslant B \Longrightarrow A' \geqslant B'$; cette orthocomplémentation définit sur T un isomorphisme dual involutif bien connu; on suppose aussi T relativement complémenté et si $A \leqslant B$, on note le complément relatif de A à B par B-A. Enfin T est complet, c'est-à-dire que tout ensemble d'éléments admet un minorant commun et un majorant commun.

II. VALUATION GENERALISEE SUR T ; ECART INDUIT SUR LES CHAINES PAR CETTE VALUATION.

DV1). Une valuation généralisée, V(A), est une application des éléments du treillis sur le corps des complexes, satisfaisant à l'axiome :

AV1). $A \leqslant B' \Longrightarrow V(A \cup B) = V(A) + V(B)$ ("valuations totales"); on en déduit que $V(\emptyset) = 0$.

ThV1). Sur une chaine de T, les nombres : $e(A,B) = |V(B-A)|$ définissent un écart, qui est dit "induit par la valuation V(A)".

III. PROBABILITE (OU LOI DE PROBABILITE), P(A), sur T.

DP1. P(A) est une valuation réelle, positive, normée, sur T.

Commentaire : La probabilité quantique se définit en spécifiant une valuation sur un treillis orthogonalisé complet comme la probabilité classique se définit en spécifiant une mesure sur un corps de Borel d'ensembles.

On peut la décrire directement par les axiomes : P(A) réel, $P(A) \geqslant 0$, $P(u) = 1$, $A \leqslant B' \Longrightarrow P(A \cup B) = P(A) + P(B)$.

Une P(A) est identique à l'écart qu'elle induit sur les chaines.

Thp1) L'équivalence de l'axiome des probabilités totales quantiques à l'axiome classique est condition nécessaire et suffisante pour que T soit booléen.

Commentaire : Sur un treillis non-booléen l'axiome quantique des probabilités totales est plus faible que l'axiome classique dont l'hypothèse, $A \cap B = \emptyset$, est impliquée par $A \leqslant B'$, mais ne l'implique pas.

ThP2). $A_i \leqslant A_j'$, pour $i \neq j$, i et j appartenant à un ensemble fini d'indices $J \Longrightarrow P(\bigcup_{i \in I} A_i) = \sum_{i \in I} P(A_i)$.

AP1). ThP2 est vrai si J est un ensemble indéfini dénombrable : σ-additivité.

ThP3). $A \leqslant B \Rightarrow P(A) \leqslant P(B)$: P(A) est non-décroissante.

IV. TREILLIS DE CERTITUDES D'UNE LOI DE PROBABILITE QUANTIQUE

ThTC1) L'ensemble des éléments, A, de T, tels que $P(A) = 1$, est fermé pour $\cup$.

ATC1) L'ensemble des éléments, A, de T, tels que $P(A) = 1$, est fermé pour $\cap$.

DTC1) Cet ensemble, qui est donc un sous-treillis de T, s'appelle "treillis de certitudes" de la loi P(A). On le note : TC(P).

ThTC2) ATC1 $\Longleftrightarrow \left[(P(A) = 0 \text{ et } P(B) = 0) \Longrightarrow (P(A \cup B) = 0)\right]$.

Commentaire : Sur un treillis de Boole, ATC1 est un théorème impliqué par la relation modulaire. Sur T, ATC1 est un axiome indépendant de ceux déjà admis.

ThTC3) Tout treillis de certitudes a un élément minimum, z(P), dit "base de la loi P(A)". (car T est complet).

ThTC4) TC(P) est identique au treillis des majorants de z(P).

ATC2) Tout élément de T autre que $\emptyset$ appartient à un treillis de certitudes.

ATC3) Tout treillis de certitudes est "déductivement complet", c'est-à-dire au sens où l'on dit qu'un système d'axiomes est complet : Un treillis de certitudes ne peut être sous-treillis d'un autre treillis de certitudes.

ThTC5) $z(P) \succ \emptyset$, ou : les bases de lois sont des atomes.

ThTC6) z(P) définit une application univoque des lois sur les atomes; c'est-à-dire : z(P) est unique.

ThTC7) T est atomique; c'est-à-dire : tout élément est minoré par un atome.

ThTC8) Les relations $Z \succ \emptyset$, $P(z) = 1$, définissent une application des atomes sur les lois.

ThTC9) Si A et B sont des unions d'atomes ; $P(A) = P(B)$, pour toute loi, implique : $A = B$.

DTC2) Si $z = z(P)$, on notera : $P(A) = P(A/z)$.

ThTC10) T satisfait à la loi de couverture : $(Z \succ \emptyset \text{ et } z \not\leqslant A) \Longrightarrow z \cup A \succ A$.

ATC4) L'application des atomes sur les lois est univoque. C'est-à-dire que, z étant donné, il n'y a qu'une loi telle que : $P(z) = 1$.

ThTC11) Il y a correspondance biunivoque entre les lois et leurs treillis de certitudes (ou entre les lois et les atomes).

Commentaire : ThTC11 n'est pas vérifié par les P classiques.
L'interprétation logique de ThTC11 paraitra satisfaisante à ceux qui pensent qu'une loi de probabilité doit être déterminée par un ensemble

de propositions constituant une donnée préalable.

On peut considérer que le "treillis de certitudes" formalise la notion classique de "catégorie d'épreuves" ou la notion quantique de "programmes de mesures".

V. PROBABILITES CONDITIONNELLES.

ThPC1) A étant un élément de T, sur le treillis orthogonal des B, minorants de A, pour lequel A est élément universel, et pour toute loi sur T, telle que $P(A/z) \neq 0$, la fonction :

$$P_1(B) = \frac{P(B/z)}{P(A/z)}$$

est une loi quantique; c'est-à-dire qu'elle satisfait à DP1.

DPC1) $P_1(B)$ est dite ; trace de la loi $P(B/z)$ sur le treillis des minorants de A.

<u>ThPC2</u>) $z \not\leq A' \Longrightarrow (z \cup A') \cap A$ est un atome.

DPC2) Pour tout atome z on pose : $(z \cup A') \cap A = E(A)\, z$ = projection orthogonale de l'atome z sur l'élément A. (c'est un atome pour $z \not\leq A'$; c'est $\emptyset$ pour $z \leq A'$).

<u>ThPC3</u>) $z \not\leq A' \Longrightarrow A = E(A)z \cup (z' \cap A)$.

ThPC4) Quelque soit l'atome z : $P(A/z) = P(E(A)z/z)$.

Commentaire : ThPC4 réduit la donnée de toutes les lois à celle d'une fonction, $P(x/y)$, des couples d'atomes.

ThPC5) Le treillis des certitudes de la loi $P_1(B)$ contient l'atome $E(A)z$.

APC1) $P_1(B)$ satisfait à ATC3.

ThPC6) $P_1(B)$ est (sur le treillis des minorants de (A) la loi déterminée par l'atome $E(A)z$: $P_1(B) = P(B/E(A)z)$.

DPC3) Nous appelons "loi conditionnée par A", et notons $P(X/A,z)$, quand la loi initiale sur T est telle que $P(A/z) \neq 0$, l'extension, au treillis total T, de la trace de la loi initiale sur le treillis des minorants de A. On a donc : $P(X/A,z) = P(X/E(A)z)$.

Commentaire : Si l'on applique la définition DPC3 en calcul classique, en utilisant, pour tout élément X du corps de Borel (ou du treillis de Boole) la décomposition (impossible sur T) : $X = (X \cap A) \cup (X \cap A')$, on trouve l'expression classique des probabilités conditionnées : $P(X/A) = P(X \cap A/A) = P(X \cap A)/P(A)$.

Mais, par suite de la détermination des lois quantiques par leur treillis de certitudes, DPC3 définit la transformation de la loi pour certaines transformations de ce treillis (ou du point de vue de la physique quantique pour certaines transformations du "programme de mesures"). Il n'existe rien d'analogue en calcul classique lors du changement de la catégorie d'épreuves.

Interprétation logique de ce qui précède

Nous interprétons l'ensemble des éléments de T comme un ensemble de propositions.

Les opérations $\cup$ et $\cap$ sur T seront interprétées comme étant respectivement la disjonction logique "ou" et la conjonction logique "et"; l'opération de complémentation sera interprétée comme étant l'opération logique de négation.

La relation d'ordre sur T doit être interprétée comme étant un prédicat de la logique (nous rappelons que l'implication logique, $a \supset b$, n'est pas un prédicat mais une opération, parfois définie par $a' \cup b$). Le prédicat logique est un prédicat "unaire", c'est-à-dire à un seul argument, pour certains auteurs ($\vdash a$ signifiant que a peut être prouvé) mais il peut être un prédicat binaire pour d'autres (13, Ch V), $a \vdash b$ s'interprétant alors par "b est déductible à partir de a", avec les propriétés :

$a \vdash a$, réflexivité, et : $(a \vdash b$ et $b \vdash c) \Longrightarrow a \vdash c$, transitivité.

Ces propriétés sont celles de la relation d'ordre sur T : $a \leqslant b$.

On s'assure que les relations fondamentales sur T entre les opérations et l'ordre sont vérifiées par les opérations logiques et la déductivité :

$(a \leqslant c$ et $b \leqslant c) \Longrightarrow a \cup b \leqslant c$, équivaut à : $(a \vdash c$ et $b \vdash c) \Longrightarrow a$ ou $b \vdash c$

$(a \leqslant b$ et $a \leqslant c) \Longrightarrow a \leqslant b \cap c$, équivaut à : $(a \vdash b$ et $a \vdash c) \Longrightarrow a \vdash b$ et c.

Un treillis de certitudes est l'ensemble des propositions déductibles de la proposition atomique z qui joue le rôle de base axiomatique d'un système déductif.

ATC3 exprime que ces systèmes déductifs sont complets au sens habituel en axiomatique.

ThTC11 exprime que l'attribution, à la base de l'un des systèmes, de la probabilité 1, c'est-à-dire l'affirmation de sa vérité, détermine la probabilité de toute proposition de l'ensemble.

Le "conditionnement quantique", par l'affirmation de la vérité de A, s'exprime par le passage de la loi déterminée par la base z à la loi déterminée par la base E(A) z

Or $E(A)z \leqslant A$ et, pour la loi initiale : $P(E(A)z/z) = P(A/z)$.

E(A)z étant le seul atome, parmi ceux qui minorent A, jouissant de cette propriété ; pour tout autre x, tel que $x \leqslant A$, on a : $P(x/z) < P(A/z)$ (inégalité stricte).

On peut donc dire que le conditionnement quantique, par A, consiste en un changement de base axiomatique, le nouvelle base étant, pour la loi initiale, la plus probable de celles desquelles on peut déduire A.

VI. RELATIONS ENTRE LA STRUCTURE DE T ET L'AXIOME DES PROBABILITES COMPOSEES

AC1) Tout élément de T est une union d'atomes.

ThC1) Pour que T soit Booléen (c'est-à-dire distributif) il faut et il suffit que les probabilités quantiques satisfassent à l'axiome des probabilités composées.

$$P(B/A,z).P(A/z) = P(A \cap B/z).$$

DC1). Pour $z \not\leqslant A'$, nous poserons :

$P(B/A,z).P(A/z) = PP(A.B/z)$ = pseudo-probabilité du couple ordonné, ou séquence, A,B. Pour $z \leqslant A'$, nous poserons : $PP(A.B/z) = 0$.

Remarque : on peut énoncer ThC1 sous la forme : Pour que T soit booléen il faut et suffit que les pseudo-probabilités de séquence soient des probabilités d'intersections : $PP(A.B/z) = P(A \cap B/z)$.

ThC2 Pour que T soit booléen il faut et il suffit que les PP satisfassent, pour tout couple, A,B, à la condition :

$z \leqslant B \Longrightarrow PP(A, B/z) + PP(A'.B/z) = 1.$

ThC3) Pour que T soit booléen, il faut et il suffit que les PP satisfassent à l'axiome quantique des probabilités totales quand le premier terme de la séquence décrit T. (D'après leur définition les PP satisfont à cet axiome quand le second terme de la séquence décrit T; la nécessité est immédiate, et ThC2 est un cas particulier de ThC3.

ThC4 Pour que T soit booléen il faut et il suffit que les PP soient fonctions symétriques des termes de la séquence : $PP(A.B/z) = PP(B.A/z)$.

Commentaire : Les PP qui sont, à la fois, une valuation et un écart sur les chaines décrites par leur second terme gardent ce caractère sur les chaines décrites par leur premier terme si, et seulement si, elles satisfont à l'axiome des probabilités composées. Le treillis T est alors booléen.

VII. AFFAIBLISSEMENT DE L'AXIOME DES PROBABILITES COMPOSEES

Commentaire : Si l'on veut préserver, pour T, la possibilité d'une structure plus faible que celle des treillis booléens, et si, cependant, on veut conserver un lien entre les pseudo-probabilités de séquences et la structure du treillis T, il faut "affaiblir" les caractères "valuation" et "écart" de ces PP sur les chaines décrites par le premier terme de la séquence. Cet affaiblissement constitue un affaiblissement de l'axiome des probabilités composées. Quant au caractère "écart", on l'affaiblira en supposant seulement une relation bi-univoque entre les PP et un écart, par exemple par le postulat suivant :

A1) La racine carrée des pseudo-probabilités de séquences est un écart sur les chaines décrites par le premier terme de la séquence : $A_1 \leqslant A' \Longrightarrow \sqrt{PP(A_1, B/z)} = e((A_1 \cup A), A)$.

ThI) $A_1 \leqslant A' \Longrightarrow |PP(A_1.B/z) + PP(A.B/z) - PP((A_1 \cup A).B/z)| \leqslant 2\sqrt{PP(A.B/z).\ PP(A_1.B/z)}$,

conséquence immédiate de la relation triangulaire, et qui, remplaçant la "relation des valuations totales", montre que AI implique un "affaiblissement" du caractère valuation des PP.

VIII. SPECIFICATION DU TREILLIS T EN VUE DE SON IMMERSION DANS (m.l.f).

AI1) L'écart dont il est question dans A1 est, quand le second terme de la séquence est un atome, induit par une valuation :

$$\sqrt{PP(A.x/y)} = |V(A.x/y)|$$

AI2) Pour trois atomes, on a : $V(y,x/z) = V(u.x/y).V(u.y/z)$. (A12 est déjà assurée en module par DPC2 et DC1; il se réduit donc à une hypothèse sur les arguments).

AI3) $V(u.x/y) = \overline{V}(u.y/x)$. (Symétrie hermitienne de la valuation).

ThI1) $P(x/y) = P(y/x)$.
(symétrie de la fonction universelle des couples d'atomes envisagée dans le commentaire de ThPC4).

AI4) Tout élément de T est une union d'atomes en nombre fini ou en infinité dénombrable : $A = \bigcup_{i \in I} x_i$, I étant un ensemble fini ou infini dénombrable d'indices.

ThI2) On peut supposer les atomes x_i de AI4 orthogonaux deux à deux; leur ensemble $\{x_i\}$ sera dit "constituer une base" de A.

DI1) $\Psi(z)$ = application des atomes, z, de T, sur le vecteur, Ψ , de l'espace euclidien complexe, H, à nombre fini ou infinité dénombrable de dimensions, défini par ses coordonnées :
$\Psi_i(z) = V(u.z_i/z)$, $\{z_i\}$ étant une base de u.

ThI3) $|\Psi_i(z)|^2 = PP(u.z_i/z) = P(z_i/z)$. D'où : $|\Psi_i(z_j)| = \delta_{ij}$

ThI4) $\sum_{i\in I} |\Psi_i(z)|^2 = P(u/z) = 1$. (Le vecteur Ψ (z) est unitaire).

ThI5) $|(\Psi(y), \Psi(z))|^2 = |\sum_{i\in I} \overline{V}(u.z_i/y).V(u.z_i/z)|^2 = P(y/z)$; (par AI3 et AI2).

DI2) $a(\Psi)$ application de tout vecteur unitaire, Ψ , de H, sur la m.l.f de dimension 1, a, que Ψ détermine.

DI3) Si $\{x_j\}$, pour $j\in J$, est une base de A, élément de T, on pose : $f(A) = \bigcup_{j\in J} f(x_j)$, qui définit l'application d'un élément quelconque, A, de T, dans (m.l.f); $f(z) = a(\Psi(z))$.

<u>ThI6)</u> L'application f(A) invarie l'ordre : $A < B \Longleftrightarrow f(A) < f(B)$.

ThI7) f(A) est indépendante de la base choisie dans A pour appliquer DI3.

<u>ThI8)</u> f(A) invarie la complémentation : $f(A') = (f(A))'$.

<u>ThI9)</u> f(A) invarie l'union : $f(A \cup B) = f(A) \cup f(B)$.

ThI10) f(A) invarie l'intersection.

ThI11 f(A) réalise l'immersion du treillis T dans le treillis (m.l.f) des multiplicités linéaires fermées d'un espace euclidien ayant, au plus, une infinité dénombrable de dimensions (Espace de Hilbert).

IX. VARIABLES ALEATOIRES ET FONCTIONS DE REPARTITION.

Les axiomes des VII et VIII ne jouent aucun rôle dans ce IX.

DVA1) Une variable aléatoire, x, à une dimension, est une application de la chaine des demi-droites $]-\infty, X[$ dans T, notée $x(]-\infty, X[)$, ou, abréviativement, x(X), et satisfaisant aux axiomes :

AVA1) Cette application invarie la structure de chaine :
$X_1 \leqslant X_2 \Longrightarrow x(X_1) \leqslant x(X_2)$.

AVA2) $x(-\infty) = \emptyset$ et $x(+\infty) = u$.

AVA3) Cette application est continue à gauche, pour toute loi :
$X' \uparrow X \Longrightarrow P(x(X')/z) \uparrow P(x(X)/z)$.

pour tout atome z.

DVA2) Fonction de répartition, f(X), de la v.a.x pour une loi quantique donnée sur T, P(A/z) :

$f(X) = P(x(X)/z)$.

ThVA1) f(X) est non-négative, non-décroissante, continue à gauche, $f(-\infty) = 0$, $f(+\infty) = 1$. C'est-à-dire f(X) est une f.r. classique.

Commentaire : Les projecteurs sur les x(X), que nous noterons E(X), sont la généralisation des "décompositions de l'unité" de M.J. von Neumann (2) auxquelles ils se réduisent quand on prend pour T le treillis (m.l.f). Les opérateurs "hypermaximaux" que ces décom-

positions de l'unité permettent de définir, pourraient être dits, à bon droit, opérateurs "à variable aléatoire liée". Or ces opérateurs ont été introduits afin d'étendre le champ de résolubilité du problème du spectre et il est tout à fait remarquable que cette extension soit solidaire de la prise en considération, pour un but purement mathématique, en dehors de toute hypothèse physique, de variables aléatoires, notion essentielle au calcul des probabilités.

DVA3) La chaine : $s(]-\infty, X[)$ ou $s(X)$, sera dite : "Chaine spectrale",

DVA4) Application des segments $[X_1, X_2[$, sur les compléments relatifs d'une chaine spectrale :
$x([X_1, X_2[) = x(]-\infty, X_2[) - x(]-\infty, X_1[) = x(X_2) - x(X_1)$.

<u>ThVA2</u>) L'application $x([X_1, X_2[)$ invarie l'intersection :
$x(([X_1, X_2[) \cap ([X_3, X_4[)) = x([X_1, X_2[) \cap x([X_3, X_4[)$.

ThVA3) L'application $x([X_1, X_2[)$ invarie la complémentation :
$x(([X_1, X_2[)') = (x([X_1, X_2[))'$.

ThVA4) L'application $x([X_1, X_2[)$ invarie l'union.

ThVA4') La continuité à gauche permet d'étendre ThVA4 à l'union dénombrable.

DVA5) Le treillis booléen défini par l'application, déduite de ce qui précède, dans T, du treillis des ensembles mesurables sur la droite réelle, sera dit : "treillis spectral" de la v.a.x, et noté $S(x)$.

DVA6) Variables aléatoires "indépendantes".

$A \in S(x)$ et $B \in S(y) \Rightarrow P(B/A, z) = P(B/z)$ et $P(A/B, z) = P(A/z)$, pour tout z.

ThVA5) Si x et y sont des v.a indépendantes, le sous treillis de T constitué par l'ensemble des éléments obtenus par union et intersection à partir de ceux de $S(x)$ et de $S(y)$ est un treillis de Boole.

Commentaire : Il n'y a donc pas de théorie proprement quantique des variables indépendantes. Les résultats classiques qui leur sont relatifs sont incorporés dans la théorie quantique; il en est ainsi, en particulier, de la théorie des fréquences et des lois des grands nombres.

C. - DÉMONSTRATIONS DE THÉORÈMES
- VÉRIFICATIONS D'AXIOMES SUR (m.l.f.)

Nous omettrons celles de ces démonstrations ou vérifications qui sont, soit immédiates, soit connues par une des théories utilisées (treillis, espace de Hilbert).

ThP1) Pour cette équivalence, il faut et suffit que l'on ait :

$A \cap B = \emptyset \Longrightarrow A < B'$, (voir commentaire), pour tout couple A, B. Supposons cette condition satisfaite :

Or : $A = (A \cap B) \cup (A - (A \cap B))$;

Mais : $[A - (A \cap B) \leqslant A$ et $(A - (A \cap B)) \cap (A \cap B) = \emptyset] \Longrightarrow (A - (A \cap B)) \cap B = \emptyset$

donc : $(A - (A \cap B)) \leqslant B'$; puis : $(A - (A \cap B)) \leqslant A \cap B'$.

d'où : $A \leqslant (A \cap B) \cup (A \cap B')$.

Mais, terme à terme : $(A \cap B) \cup (A \cap B') \leqslant A$;

Donc : $A = (A \cap B) \cup (A \cap B')$; et, par dualité : $A' = (A' \cup B') \cap (A' \cup B)$.

Ceci est un cas particulier de distributivité, qui implique la distributivité générale; En effet :

posons : $X = A \cap (B \cup C)$, et : $Y = (A \cap B) \cup (A \cap C)$; terme à terme : $Y \leqslant X$.

Or : $X - Y \leqslant X$ et $X - Y \leqslant Y' = (A' \cup B') \cap (A' \cup C')$;

D'où : $X - Y \leqslant A \cap (B \cup C) \cap (A' \cup B') \cap (A' \cup C')$; si, donc : $A = (A \cup B) \cap (A \cup B')$, et $B' = (A \cup B') \cap (A' \cup B')$ il vient :

$X - Y \leqslant (A \cup B) \cap (B \cup C) \cap B' \cap (A' \cup C') \leqslant B'$; et de même : $X - Y \leqslant C'$;

d'où : $X - Y \leqslant B' \cap C' = (B \cup C)'$; or $X - Y \leqslant B \cup C$; donc : $X - Y = \emptyset$.

C'est-à-dire : $A \cap (B \cup C) = (A \cap B)(A \cap C)$.

L'hypothèse d'équivalence des axiomes classique et quantique implique donc la distributivité de T ; T est alors distributif et complémenté, c'est-à-dire Booléen. Réciproquement, si T est booléen, sa distributivité implique immédiatement que $A \cap B = \emptyset \Longrightarrow A \leqslant B'$, c'est-à-dire l'équivalence des axiomes.

ThTC6) $z_1 = z(P)$ et $z_2 = z(P)$ et $z_1 \neq z_2 \Longrightarrow P(z_1) = 1$ et $P(z_2) = 1$

$\Longrightarrow P(z_1 \cap z_2) = 1 \Longrightarrow P(\emptyset) = 1$ ce qui est absurde. Donc : $z_1 = z_2$.

ThTC10) Démontrons que, si $z \succ \emptyset$ et $z \not\leqslant A$, l'existence d'une chaine : $A < B < A \cup z$, est contradictoire (inégalités strictes).

On aurait alors : $A \cup z = A \cup (B - A) \cup ((A \cup z) - B)$, les trois termes de cette union étant, deux à deux, orthogonaux.

donc : $P(A \cup z) = P(A) + P(B - A) + P((A \cup z) - B) = P(A) + P((B - A) \cup ((A \cup z) - B))$.

Soit x un atome appartenant à $(B - A) \cup ((A \cup z) - B)$. Pour la loi déterminée par x le dernier terme des relations précédentes est égal à 1; il est donc impossible que $P(z/x) = 0$, sinon, par ThTC2 le premier terme serait nul. Donc, par ThTC7, on a :

$((B - A) \cup ((A \cup z) - B)) \cap z' = \emptyset$; donc : $((B - A) \cup ((A \cup z) - B)) \cup z' > z'$,

L'inégalité étant stricte. Mais $u \succ z'$; donc : $((B - A) \cup ((A \cup z) - B)) \cup z' = u$, et : $((B - A) \cup ((A \cup z) - B)) \cup z' \succ z'$; donc : $((B - A) \cup ((A \cup z) - B))$ est un atome, ce qui est contradictoire avec le fait qu'il est une union de deux termes orthogonaux.

ThPC2) $z \cup A' \succ A' \Longrightarrow (z \cup A') \cap A \succ A' \cap A = \emptyset$.

ThPC3) La relation : $A = E(A)z \cup (z' \cap A)$ est immédiate si $E(A)z = A$.

Si $E(A)z \neq A$, posons : $A - E(A)z = B$. On a : $B \propto A$. Mais, de $z \cup A' \succ A'$, on déduit $z' \cap A \propto A$. Donc : $B = z' \cap A$, ou bien $A = B \cup (z' \cap A)$. Or la seconde éventualité est exclue car : $B \leqslant (E(A)z)' = (z' \cap A) \cup A' \Longrightarrow B \cup (z' \cap A) \leqslant (z' \cap A) \cup A'$, ce qui, si cette seconde éventualité était vraie, deviendrait : $A \leqslant (z' \cap A) \cup A' \Longleftrightarrow (z' \cap A) \cup A' = ((z' \cap A) \cup A') \cup A = u \Longleftrightarrow (z \cup A') \cap A = \emptyset$, ce qui est exclu. Donc c'est la première éventualité qui est vraie : $B = z' \cap A \Longrightarrow A = E(A)z \cup (z' \cap A)$. ThPC3 reste vrai si $z \leqslant A'$.

ThC1) La condition est nécessaire : Sur un treillis booléen :

$B = (A \cap B) \cup (A' \cap B)$. d'où : $P(B/A,z) = P(A \cap B/A,z) + P(A' \cap B/A,z) = P(A \cap B/E(A)z) + P(A' \cap B/E(A)z) = P(A \cap B/E(A)z)$, et, comme $A \cap B$ appartient au treillis des minorants de A, par ThPC6 : $P(A \cap B/E(A)z) = P_1(A \cap B) = P(A \cap B/z)/P(A/z)$, c.q.f.d.

La condition est suffisante :

si elle est satisfaite, on a ; pour tout couple A, B : $P((A \cap B) \cup (A \cap B')/z) = P(A \cap B/z) + P(A \cap B'/z) = (P(B/A,z) + P(B'/A,z))P(A/z) = P(A/z)$,

pour tout z. Donc, par AC1 et ThTC9 : $(A \cap B) \cup (A \cap B') = A$. Ce qui implique la distributivité, comme on l'a vu au ThP1.

ThC2) Si T est booléen, par ThC1 :

$PP(A.B/z) + PP(A'.B/z) = P(A \cap B/z) + P(A' \cap B/z) = P(B/z) = 1$, si $z \leqslant B$.

Réciproquement :

$z \leqslant B \Rightarrow P(B/E(A)z)P(A/z) + P(B/E(A')z)P(A'/z) = 1 = P(A/z) + P(A'/z)$

$\Longleftrightarrow P(B'/E(A)z)P(A/z) + P(B'/E(A')z)P(A'/z) = 0$

$\Longleftrightarrow P(B'/E(A)z)P(A/z) = 0$ et $P(B'/E(A')z)P(A'/z) = 0$

$\Longleftrightarrow (z \leqslant A'$ ou $E(A)\,z \leqslant B)$ et $(z \leqslant A$ ou $E(A')\,z \leqslant B)$

$\Longleftrightarrow (z \leqslant A$ et $E(A)z \leqslant B)$ ou $(E(A')z \leqslant B$ et $z \leqslant A')$ ou $(E(A)z \leqslant B$ et $E(A')z \leqslant B)$

$\Rightarrow (z \leqslant A \cap B)$ ou $(z \leqslant A' \cap B)$ ou $((E(A)z \cup E(A')z) \leqslant (A \cap B) \cup (A' \cap B))$.

Mais : $z \leqslant E(A)z \cup E(A')z$, par ThPC4. D'où l'on déduit que chacun des termes de la dernière disjonction triple implique :

$z \leqslant (A \cap B) \cup (A' \cap B)$, pour tout $z \leqslant B$; donc, par AC1 :

$B \leqslant (A \cap B) \cup (A' \cap B)$, d'où la distributivité suit comme plus haut (ThP1).

ThC4) La condition : $PP(A.B/z) = PP(B.A/z)$, est nécessaire pour que T soit booléen, puisque, dans ce cas, d'après ThC1, chacun des termes est égal à $P(A \cap B/z)$.

Réciproquement, si $PP(A.B/z) = PP(B.A/z)$, le ThC2 s'applique immédiatement par le second membre.

ThI6) Si z est un atome de T, f(z) est un atome de (m.l.f) et l'on a :

$P(A/z) = \sum_{j \in J} P(X_j/z) = \sum_{j \in J} P(f(x_j)/f(z)) = P(f(A)/f(z))$.

D'où $z \leqslant A \Longleftrightarrow P(A/z) = 1 \Longleftrightarrow f(z) \leqslant f(A)$.

Or : $A \leqslant B \Longleftrightarrow x_j \leqslant B$, pour tout $j \Longleftrightarrow f(x_j) \leqslant f(B) \Longleftrightarrow f(A) \leqslant f(B)$.

ThI8) donc : $z \leqslant A' \Longleftrightarrow f(z) \leqslant f(A')$.

Mais d'autre part ; $z \leqslant A' \Longleftrightarrow P(A/z) = 0 \Longleftrightarrow f(z) \leqslant (f(A))'$.

d'où : $f(A') = (f(A))'$.

ThI9) $f(z) \leqslant (f(A \cup B))' \Longleftrightarrow f(z) \leqslant f((A \cup B)') \Longleftrightarrow z \leqslant (A \cup B)' = A' \cap B'$

$\Longleftrightarrow z \leqslant A'$ et $z \leqslant B' \Longleftrightarrow f(z) \leqslant f(A')$ et $f(z) \leqslant f(B')$

$\Longleftrightarrow f(z) \leqslant (f(A))'$ et $f(z) \leqslant (f(B))' \Longleftrightarrow f(z) \leqslant (f(A))' \cap (f(B))'$

$\Longleftrightarrow f(z) \leqslant (f(A) \cup f(B))'$.

D'où : $(f(A \cup B))' = (f(A) \cup f(B))'$;

$\Longleftrightarrow f(A \cup B) = f(A) \cup f(B)$.

ThVA2) On a : $P(x([X_1, X_2[)) = P(x(X_2)) - P(x(X_1)) = P(]-\infty, X_2[) - P(]-\infty, X_1[)$
$= P([X_1, X_2[)$ pour la loi qui admet f(X) pour f.r. sur l'axe réel des X.

Nous noterons schématiquement la relation à démontrer : $x(12 \cap 34) = x(12) \cap x(34)$.

Supposons d'abord que : $12 \cap 34 \neq \emptyset$ (segments non disjoints).
$z \leqslant x(12 \cap 34) \Longleftrightarrow P(x(12 \cap 34)/z) = 1 \Longleftrightarrow P(12 \cap 34) = 1 \Longleftrightarrow P(12) = 1$ et

$P(34) = 1 \Longleftrightarrow P(x(12)) = 1$ et $P(x(34)) = 1 \Longleftrightarrow P(x(12) \cap x(34)) = 1$ (par ATC1) $\Longleftrightarrow z \leqslant x(12) \cap x(34)$; d'où, par AC1 et ThTC9 : $x(12 \cap 34) = x(12) \cap x(34)$.

D'autre part : $12 \cap 34 = \emptyset \Longrightarrow P(12 \cap 34) = 0$ pour toute loi $\Longleftrightarrow$ $P(x(12 \cap 34)/z) = 0$ pour tout $z \Longleftrightarrow x(12 \cap 34) = \emptyset$ (par ATC2), ce qui complète la preuve.

ThVA5) DVA6 $\Longrightarrow PP(A.B/z) + PP(A'.B/z) = P(B/z) \Longrightarrow$ (par ThC2).

$B = (A \cap B) \cup (A' \cap B)$ qui implique, comme en la preuve de ThP1, la distributivité générale.

Remarque sur A1)

Une forme générale de l'affaiblissement du caractère "valuation" des PP quand le premier terme de la séquence décrit une chaine de T est :

$$A_1 \leqslant A' \Longrightarrow \left| PP(A_1.B/z) + PP(A.B/z) - PP((A_1 \cup A).B/z) \right| \leqslant$$
$$\leqslant f(PP(A_1.B/z), PP(A.B/z)), \quad (1)$$

Une forme générale de l'affaiblissement du caractère "écart" des PP dans les mêmes conditions est :

$$A_1 \leqslant A' \Rightarrow e((A_1 \cup A), A) = g(PP(A_1.B/z)).$$

qui, par la relation triangulaire implique :

$$\left| g^2(PP(A_1.B/z)) + g^2(PP(A.B/z) - g^2(PP((A_1 \cup A).B/z)) \right| \leqslant$$
$$\leqslant 2\sqrt{g^2(PP(A_1.B/z)).g^2(PP(A.B/z))}. \quad (2)$$

Or (1) et (2) sont des conditions de même forme et la limitation imposée au choix des PP sera minimum si elles sont équivalentes. L'axiome AI assure cette équivalence.

Vérification de l'axiome, AI1 sur (m.l.f).

On sait qu'on pose : $P(A/z) = \| E(A).\Psi(z) \|^2$, où : $\Psi(z)$ est un vecteur unitaire appartenant à la m.l.f de dimension un, z.

Et ensuite : $P(B/A, z) = \| E(B).E(A).\Psi(z) \|^2 / \| E(B).\Psi(z) \|^2$

d'où : $PP(A.B/z) = \| E(B).E(A).\Psi(z) \|^2$.

Par suite, pour deux atomes, x et y :

$PP(A.x/y) = \| E(x).E(A).\Psi(y) \|^2 = (E(x).E(A).\Psi(y), E(A).\Psi(y))$

mais : $E(x).E(A).\Psi(y) = k.\Psi(x)$;

d'où : $k = (E(x).E(A).\Psi(y), \Psi(x)) = (E(A).\Psi(y), \Psi(x))$.

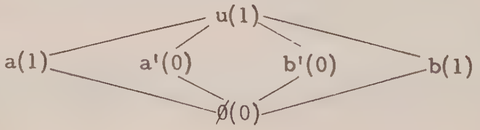
D'où :

$PP(A.x/y) = (E(A).\Psi(y), \Psi(x)).(\Psi(x), E(A).\Psi(y)) = |(E(A).\Psi(y), \Psi(x))|^2$

Or le produit scalaire $(E(A).\Psi(y), \Psi(x))$ satisfait à DV1 et à AV1 : donc AI1 est vérifié ; AI2 et AI3 sont ensuite immédiatement vérifiés.

Indépendance de ATC1 et des axiomes qui le précèdent :

Sur le treillis ci-contre, supportant une P dont les valeurs sont données à côté de chaque élément, les axiomes précédant ATC1 sont vérifiés, mais ATC1 ne l'est pas.



RELATIONS ENTRE LE PRESENT MEMOIRE ET CERTAINS TRAVAUX ANTERIEURS :

La relation entre les pseudo-probabilités de séquences et les écarts sur les chaines de T peut s'écrire, quand $A_1 \leqslant A_2$:

$$e(A_1, A_2) = \sqrt{PP((A_2 - A_1).B/z)} = \| E(B).E(A_2 - A_1).\Psi(z) \|$$

Supposons que A_1 et A_2 soient deux éléments de la chaine spectrale d'une v.a particulière : la variable dynamique q (coordonnée généralisée) :

$$A_1 = q(]-\infty, Q_1[) = q(Q_1), \text{ et, } A_2 = q(]-\infty, Q_2[) = q(Q_2), \text{ avec } Q_1 \leqslant Q_2.$$

Des propriétés de l'application q découle immédiatement que $e(A_1, A_2)$ est <u>aussi</u> un écart sur l'axe réel des Q :

$$e(A_1, A_2) = e(Q_1, Q_2).$$

Supposons que B soit une m.l.f. spectrale de la v.a. particulière p, moment conjugué de q ; on sait que q et p ne commutent pas.

J'appelle "séquence immédiate" de propositions expérimentales la suite ordonnée : $(A_2 - A_1)$ puis $(B_2 - B_1)$, où : "un corpuscule <u>était</u> sur $[Q_1, Q_2[$ dont l'impulsion <u>est</u> sur $[P_1, P_2[$". L'expression séquence immédiate est destinée à rappeler que l'on suppose qu'il n'y a pas évolution temporelle de l'état $\Psi(z)$.

Laissant P_1 et P_2 fixés, considérons l'ensemble des séquences immédiates obtenues en faisant varier Q_1 et Q_2, c'est-à-dire, l'ensemble des attributions <u>a posteriori</u> d'une position à un corpuscule d'impulsion connue.

Le formalisme fait correspondre à cet ensemble, pour un état $\Psi(z)$ donné, un écart sur l'axe des Q : $e(Q_1, Q_2)$. En ce sens il donne un contenu effectivement topologique aux propositions de localisation a posteriori. L'opinion, courante parmi les physiciens de l'école de Copenhague, que de telles propositions sont vides de sens, n'est pas fondée.

Dans une conférence que nous avons faite en 1953 au séminaire de M. le Professeur Darmois nous avions construit le formalisme quantique dans le cas particulier de deux variables dynamiques conjuguées en réglant axiomatiquement cette localisation a posteriori. le mémoire actuel englobe leur cas dans une théorie générale.

REMARQUES

REMARQUE sur l'identification des relations : $a \vdash b$ et $a \leqslant b$:

En calcul des propositions : $(a \vdash b) \Longrightarrow (\vdash a \supset b)$, c'est un cas particulier du théorème de déduction (13, § 21);

et : $(\vdash a \supset b) \Longleftrightarrow (\vdash a' \cup b) \Longleftrightarrow P(a' \cup b) = 1$, pour toute loi de probabilité à deux valeurs 0 et 1 (13, § 28, § 29) or sur un treillis ayant un élément universel, u, seul cet élément satisfait à la condition : $P(u) = 1$, pour toute loi; donc :

$$P(a' \cup b) \equiv 1 \Longleftrightarrow a' \cup b = u.$$

Sur tout treillis : $(a \leqslant b) \Longrightarrow (a' \cup b = u)$, la réciproque n'étant vraie que si ce treillis est booléen.

Sur un treillis booléen on a donc :

$$(a \vdash b) \Longrightarrow (\vdash a \supset b) \Longleftrightarrow (a \leqslant b).$$

et l'identification de $a \vdash b$ et $a \leqslant b$ constitue un affaiblissement de $a \vdash b$.

Mais, sur un treillis non distributif, $a \vdash b$ et $a \leqslant b$ sont, toutes deux, plus fortes que $\vdash a \supset b$; en ce sens, leur identification est plus "acceptable" sur un treillis non distributif tel que (m.l.f.) que sur un treillis de Boole.

REMARQUE

- Relations entre les postulats qui règlent un calcul des probabilités sur un treillis et la modularité ou la distributivité de ce treillis :

Nous voulons, en particulier, préciser les liens entre la "relation modulaire" :

$$P(x) + P(y) = P(x \quad y) + P(x \cap y) \quad , \quad (RM)$$

d'une part, et la modularité : $A \leqslant C \Longrightarrow A \cup (B \cap C) = (A \cup B) \cap C$, (M), ou la distributivité : $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, (D) d'autre part.

Dans le mémoire de von-Neumann et Birkhoff, on postule (RM), non pour une probabilité, mais pour une "fonction de dimension", $d(x)$, qui est, de plus, supposée strictement croissante :

$$x < y \Longrightarrow d(x) < d(y).$$

Or : $A \cup (B \cap C) \leqslant (A \cup B) \cap C$ (terme à terme) ; mais :

$$(RM) \Longrightarrow d(A \cup (B \cap C)) = d((A \cup B) \cap C),$$

qui exclut l'inégalité stricte ; d'où (M).

Une fonction de probabilité n'est pas strictement croissante, mais supposons qu'elle satisfasse à ATC2 (tout élément de T, autre que $\emptyset$, appartient à un treillis de certitudes) ; Si on lui impose (RM) on en déduit AVA1, puis :

$$p\left[(A \cup B) \cap C) - (A \cup (B \cap C))\right] = 0, \text{ d'où } (M).$$

Donc : $(ATC2 \text{ et } RM) \Longrightarrow (M)$.

Supposons maintenant que l'on impose au calcul des probabilités, en plus de ATC2 et de RM, le postulat AC1 (tout élément de T est une union d'atomes). Alors, par RM :

$x \cap y = \emptyset \Longrightarrow P(x) + P(y) = P(x \cup y)$; puis, par AVA1 (impliqué par RM) : $P(x) = P(x \cup y - y)$; puis, par AC1 et ThTC9 : $x = x \cup y - y$ ce qui implique : $x \leqslant y'$; en résumé : $x \cap y = \emptyset \Longrightarrow x \leqslant y'$; il y a donc équivalence de l'axiome quantique et de l'axiome classique des probabilités totales ; d'où, par ThP1, la distributivité (D) : $(ATC2 \text{ et } RM \text{ et } AC1) \Longrightarrow (D)$.

Donc, sur (m.l.f.), dans un espace de Hilbert de dimension finie, si l'on suppose ATC2 et AC1, il faut abandonner RM, puisque un tel treillis n'est pas distributif ; et, sur (m.l.f), dans un espace de Hilbert de dimension infinie, si l'on suppose simplement ATC2, il faut abandonner RM, puisqu'un tel treillis n'est même pas modulaire.

RÉFÉRENCES

A) Sur l'objet de ce mémoire :

(1) J. von Neumann et G. Birkhoff. The logic of quantum mechanics. Annals of math. 37. (1936).

(2) J. von Neumann. Les fondements mathématiques de la mécanique quantique (Trad. Proca. 1946).

(3) L. de Broglie. La statistique des cas purs en mécanique ondulatoire et l'interférence des probabilités (Revue Scientifique 87. 1948).

4). R. Fortet : Faut-il élargir les axiomes du calcul des probabilités? (Congrès intern. de philo. des Sc. 1949, Actualités Sc. Hermann n° 1146).

5). Actes du 2ème colloque intern. de Logique math. (Collection de Log. math. A, V, Gauthier-Villars, 1954). (Mme P. Février, Mrs Reichenbach, J.L. Destouches, Bodiou ...).

6). Mme P. Destouches-Février : La structure des Théories physiques (1951).

7). G. Bodiou : Conférence au Séminaire de M. le Professeur G. Darmois, 22/5/1953 (ronéotypée ; publiée dans : Annales de la Fac. des Sc. de Marseille, XXII, 2, 1953).

b). Recherches sur les fondements du calcul quantique des probabilités dans les cas purs (Thèse. Annales de Physique 1950).

c). Relations entre la seconde quantification, la mécanique ondulatoire dans l'espace de configuration et les problèmes du second ordre en calcul des probabilités. (Journal de Phys. 15. 1954).

B). Ouvrages généraux :

8). A.N. Kolmogoroff : Foundations of the theory of Probability (Transl. by N. Morrison, 2d ed. 1956).

9). G. Birkhoff : Theory of lattices (1940-48).

10). Mme M.L. Dubreil-Jacotin, L. Lesieur, R. Croisot : Théorie des treillis, des structures algébriques ordonnées, et des treillis géométriques (1953).

11). G. Julia : Introduct. math. aux théories quantiques. (II, 1938).

12). F. Riesz et B. Sz. Nagy : Leçons d'analyse fonctionnelle (1953-55).

13). S.C. Kleene : Introduction to metamathematics (1952).

14). R. Carnap : Logical foundations of probability (1950).

SUR LES AGRÊGATS DE DISTRIBUTIONS STATISTIQUES D'UNE MÊME FAMILLE

par

P. THIONET

PRÉAMBULE

1 - La notion de distribution statistique est bien connue. Par exemple la distribution des revenus (des habitants d'un pays donné) s'obtient en comptant combien d'habitants ont des revenus supérieurs à 5 millions, compris entre 4 et 5 millions, entre 3 et 4 millions, etc ... "Distribution des habitants suivant leurs revenus" serait plus correct.

Distribution et répartition sont synonymes.

On peut envisager encore la distribution des exploitations agricoles suivant leur superficie, la distribution des établissements industriels et commerciaux suivant leur chiffre d'affaires, la distribution des mêmes établissements suivant le nombre de leurs salariés, etc ...

2 - On définit par ailleurs des lois théoriques de distribution :

Loi de Gauss (pour une variable pouvant être négative),
Loi de Galton,
Loi de Pareto,
etc ...

Les distributions statistiques courantes ont souvent une ressemblance marquée avec telle ou telle loi théorique de distribution : par exemple certaines distributions de revenus imposables sont assez exactement superposables à une loi de Pareto ("ajustée" convenablement).

3 - Quand on réunit 2 populations dont les distributions (de revenus, pour fixer les idées) suivent chacune une certaine loi théorique, il est bien connu que la distribution résultante n'a pas en général la même allure que les distributions composantes. Par exemple si les composantes sont unimodales (et de modes différents) la résultante présente très souvent 2 modes, dont chacun est voisin du mode des composantes.

En langage plus familier, la réunion de 2 distributions à une bosse donne le plus fréquemment une distribution à 2 bosses.

4 - Nous conviendrons d'employer ici les termes suivants :

- les distributions composantes seront des distributions de base;
- en réunissant deux populations de base, on obtient un agrégat.

Ainsi par exemple si l'on mélange 2 distributions de Galton, l'agrégat n'est plus une distribution de Galton quand les modes des distributions de base sont distincts. Lorsque les modes sont confondus, on peut encore a-

avoir des doutes; et le calcul montre bien que l'agrégat n'est pas une distribution de Galton (même s'il existe une certaine ressemblance graphique entre les deux).

Le calcul montre de même que des distributions de base de Pareto ne donnent pas d'agrégat qui suive une loi de Pareto.

5 - On s'est donc posé la question suivante : Existe-t-il des types de distribution assez généraux (ou assez particuliers) pour qu'ils puissent représenter à la fois :
- certaines populations de base,
- les agrégats de celles-ci?

C'est un problème de mathématiques pures.

6 - Pourtant les origines de ce problème sont d'ordre statistique. Par exemple, M. Fonsagrive, notre collègue à l'I.N.S.E.E., s'est posé cette question, à propos d'une étude de M. Marchand, publiée au Bulletin de la Statistique Générale de la France de Janvier-Mars 1945 et intitulée : La concentration du personnel dans les entreprises en France entre les deux guerres. On expose dans cette étude les résultats d'un ajustement de lois de Galton sur les distributions des entreprises suivant leur nombre de salariés (à l'intérieur des grandes branches d'activités collectives). Beaucoup plus tard M. Fonsagrive a attiré notre attention sur ce problème. Des problèmes d'agrégation se posent constamment en économétrie (théorie des choix, courbes d'Engel, élasticité de la demande, etc.) et ont fait l'objet de travaux récents de MM. Nataf, Malinvaud, en France.

7 - Nous avons eu l'idée d'appliquer au problème de l'agrégation des distributions statistiques la méthode de la courbe de concentration de C. Gini dont nous avions déjà fait usage pour résoudre en 1954 un problème de sondage, à nous posé par la Direction Générale des Impôts (1).

Une première étape des recherches, très brève et presque d'ordre expérimental, n'a porté que sur quelques courbes de concentration particulières, rencontrées au cours du travail sur les sondages. L'énoncé du problème et les résultats fragmentaires alors trouvés (3 exemples traités) (qui formaient une première note de mai 1955) constitueront la 1ère partie du présent exposé, avec quelques rappels de résultats concernant la Courbe de Concentration de C. Gini et Lorenz.

8 - En décembre 1955, M. Fréchet, nous fit savoir (comme suite à notre note) qu'il avait établi un théorème, d'après lequel n'importe quelle courbe de concentration possède la propriété que nous avions constatée sur le premier de nos 3 exemples.

Par la suite, nous avons donné du théorème de M. Fréchet une démonstration un peu nouvelle puis (encouragé par M. Fréchet à poursuivre ces recherches) nous avons étendu cette démonstration à des cas beaucoup plus généraux, correspondant à nos Exemples 2 et 3, car le théorème initial concernait seulement une famille de distributions ne différant entre elles que par les deux paramètres : effectif de la population et valeur moyenne de la variable.

Nous sommes arrivé à l'énoncé suivant :
(Etant donné une famille quelconque de distributions, d'une variable
($x(-\infty < a \leqslant x \leqslant b)$, à un nombre quelconque de paramètres ; pour que l'agrégat
(de telles distributions de base fasse lui-même partie de la famille, il faut
(que les moyennes arithmétiques des distributions de base soient égales
(entre elles.

---

(1) Voir Journ. Sté Stat. Paris. Juillet-Sept. 1955, page 192.

La condition n'est pas suffisante; mais nous donnerons la forme générale des familles de distributions possédant cette propriété (elles dépendent de n fonctions arbitraires et n-1 paramètres, outre la moyenne commune des distributions).

9 - Ainsi il n'existe pas, par exemple, de loi unimodale telle qu'une famille de distributions ayant même mode (et des moyennes différentes) puisse comprendre en son sein ses propres agrégats.

10 - Les démonstrations du théorème de M. Fréchet et de son extension font l'objet de la seconde partie du présent exposé.

Nous tenons à exprimer ici notre gratitude à M. Fréchet et à M. Malinvaud pour le concours décisif qu'ils nous ont apporté dans cette étude.

PREMIÈRE PARTIE

POINT DE DÉPART ET PREMIÈRES RECHERCHES

I. - DÉFINITION DES LOIS DE DISTRIBUTION PAR LEUR COURBE DE CONCENTRATION

1 - LA COURBE DE CONCENTRATION DE LORENZ GINI

Rappelons la définition de la courbe de concentration (qu'on rencontre chez les auteurs italiens, plus que chez les auteurs anglo-saxons ou français) :

Considérons la distribution d'une variable positive x (par exemple le chiffre d'affaires d'un établissement industriel et commercial) ; on introduit 2 variables p et q comprises entre 0 et 1; à savoir :

p, proportion des redevables dont le C.A. est inférieur à x ;

q, proportion du total des C.A. correspondant aux redevables dont le C.A. est inférieur à x.

Par définition, le lieu du point de coordonnées cartésiennes (p,q) est la courbe de concentration; soit :

$$g(p,q) = 0$$

l'équation de cette courbe (1). On a évidemment :

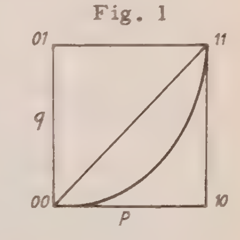
$$g(00) = g(11) = 0$$

ainsi qu'une condition de croissance $\frac{dq}{dp} > 0$

c'est-à-dire $g'_p \; g'_q < 0$

et une condition de convexité $\frac{d^2 q}{dp^2} > 0$

Fig. 1



Remarque : La notion de courbe de concentration s'étend à une variable pouvant présenter des valeurs négatives, par exemple les bénéfices des entreprises (industrielles et commerciales), le total des bénéfices restant positif.

(Voir Fig. 2 la forme de la courbe de concentration).

En pareil cas la condition de croissance ci-dessus disparait; celle de convexité subsiste.

Nous ne ferons guère usage de cette extension; mais nous admettrons une fois pour toutes que la variable x est comprise entre a et b, avec $(-\alpha < a)$

2 - LOI DE PROBABILITE ET COURBE DE CONCENTRATION

On passe de la loi de probabilité à la courbe de concentration par les formules suivantes :

Soit : $f(x)$: la densité de probabilité (dont on suppose l'existence);

: m : l'effectif total de la population (d'ailleurs infini);

: $m\bar{x}$: (x) le C.A. total de la dite population.

On voit facilement que :

$$\left| \begin{array}{l} m f(x) = m \dfrac{dp}{dx} \\ x = \bar{x} \dfrac{dq}{dp} \end{array} \right\} \quad \text{d'où } f(x) = \frac{1}{\bar{x}} \cdot \frac{1}{\dfrac{d^2q}{dp^2}}$$

Fig. 2

3 - EXEMPLE : LOI DE PARETO : $p = 1 - \left(\frac{a}{x}\right)^{\alpha}$

où a désigne le C.A. minimum et où $1,6 < \alpha < 1,9$. Il vient d'une part :

$$m f(x) = \alpha \frac{m}{a} \left(\frac{a}{x}\right)^{\alpha+1}$$

d'autre part :

$$dq = \frac{1}{\bar{x}} \frac{x f(x)}{m} dx = \frac{\alpha}{x} \left(\frac{a}{x}\right)^{\alpha} dx$$

d'où

$$q = \frac{\alpha}{\alpha - 1} \frac{a}{\bar{x}} \left[1 - \left(\frac{a}{x}\right)^{\alpha-1} \right] \quad (\text{avec } \alpha > 1)$$

Mais

$$\bar{x} = \int_a^{+\infty} x f(x)\, dx = \frac{\alpha}{\alpha - 1} a$$

d'où : .

$$q = 1 - \left(\frac{a}{x}\right)^{\alpha-1}$$

D'où

$$g(p,q) = (1 - p)^{\alpha-1} - (1 - q)^{\alpha} = 0$$

4 - DISTRIBUTIONS DEFINIES PAR L'EQUATION DE LEUR COURBE DE CONCENTRATION $g(p,q) = 0$

a - Si l'on résoud en q l'équation de la courbe de concentration, x et $f(x)$ s'en déduisent (par 2 dérivations), en fonction du paramètre arbitraire $\bar{x}$. Ainsi la courbe de concentration définit une certaine famille de distributions statistiques "semblables entre elles"; la valeur moyenne de la variable caractérise chacune des courbes de cette famille.

Il est remarquable que les courbes de concentration les plus simples et les plus naturelles ne correspondent pas à des distributions d'usage courant. En voici quelques exemples.

b - Nous nous sommes attardé sur le cas où cette courbe serait assimilable à un arc de conique, c'est-à-dire :

$$g(p,q) = a\,p\,(1-p) + b\,q\,(1-q) + c\,p\,(1-q) + d\,q\,(1-p)$$
$$= (a+c)\,p + (b+d)\,q - \left[a\,p^2 + (c+d)\,pq + bq^2\right]$$

Et nous avons trouvé que cet arc présentait la forme voulue, à condition d'avoir simultanément :

$$(a+c)(b+d) < 0$$
$$(b+c)(a+b+c+d) < 0$$
$$(b+d)(ab-cd)(a+b+c+d) > 0$$

c - La courbe ainsi définie s'est révélée trop lourde pour la pratique. Nous sommes passé de 3 à 1 paramètre, en faisant usage de l'hyperbole équilatère d'équation :

$$p\,q + \varepsilon\,p - (1+\varepsilon)\,q = 0 \qquad (\varepsilon > 0)$$

qui présente d'ailleurs le grave défaut d'être symétrique par rapport à la droite d'équation

$$p + q = 1$$

ce qui n'est pas le cas de beaucoup de courbes de concentration réelles. On obtient :

$$x = \frac{\overline{x}\,\varepsilon(1+\varepsilon)}{(1+\varepsilon-p)^2}$$
$$f(x) = \left[\frac{m}{2}\sqrt{\overline{x}\,\varepsilon(1+\varepsilon)}\right]\frac{1}{x^{3/2}}$$

Fig. 3

II. - LE PROBLÈME DE L'AGRÉGATION DES LOIS DE DISTRIBUTION

1 - AGREGATION DE PLUSIEURS DISTRIBUTIONS

a) Considérons par exemple la distribution des C.A. des épiciers, agrégat des deux distributions relatives aux épiciers détaillants et aux épiciers en gros. Soit

(1) $$g(p,q) = 0 \quad , \quad g_1(p_1\,q_1) = 0 \quad ,$$

les équations des 2 courbes de concentration correspondantes. On admet que les C.A. des deux distributions sont exprimés dans la même monnaie; la variable x est donc commune aux 2 distributions et à leur agrégat; et il vient :

(2) $$\overline{x}\,\frac{dq}{dp} = \overline{x}_1\,\frac{dq_1}{dp_1}$$

ou encore

(2)' $$\frac{\frac{1}{m}\frac{\partial g}{\partial p}}{\frac{1}{(x)}\frac{\partial g}{\partial q}} = \frac{\frac{1}{m_1}\frac{\partial g}{\partial p}}{\frac{1}{(x_1)}\frac{\partial g}{\partial q}}$$

Fig. 4

Cette relation entre les pentes des 2 courbes de concentration définit une correspondance entre points homologues M et M_1 correspondant à une même valeur de la variable x.

b - Soit P, Q les coordonnées d'un point de la courbe de concentration de l'agrégat. On a les relations suivantes :

$$\begin{array}{ll} \text{entre les effectifs} & : m + m_1 = M \\ \text{entre les C.A. totaux} & : (x) + (x_1) = (X) \end{array} \Bigg\} (3) \quad \begin{array}{l} : m\,p + m_1\,p_1 = M\,P \\ : (x)\,q + (x_1)\,q_1 = (X)\,Q \end{array} \Bigg\} (4)$$

L'équation de la courbe de concentration de l'agrégat s'obtiendra en éliminant les 4 quantités $p\ q\ p_1\ q_1$ entre les 5 relations (1) (2) et (4).

c - On étend immédiatement cet énoncé au cas d'un agrégat de plus de 2 distributions.

2 - SUR LES FAMILLES DE DISTRIBUTIONS SUSCEPTIBLES D'AGREGATION.

Divers auteurs ont proposé des expressions (dépendant de certains paramètres) susceptibles de représenter (plus ou moins heureusement) la distribution des revenus ou des salaires, des effectifs, des chiffres d'affaires, etc ... d'un pays donné (lois de Pareto, de Galton, etc ...). Ils ne semblent pas s'être posé la question suivante : la distribution statistique (de revenus, de salaires, etc ...) ainsi considérée résulte de la juxtaposition de distributions analogues concernant des sous-populations. Par exemple :

La distribution des établissements industriels et commerciaux français suivant leur chiffre d'affaires résulte de la juxtaposition des distributions des chiffres d'affaires des boucheries, des épiceries, des fonderies, des banques, etc ...

On se trouve donc devant le problème suivant : existe-t-il des lois de distribution assez générales pour représenter la distribution des <u>chiffres d'affaires d'un secteur économique</u>?

- étant entendu qu'en agrégeant les distributions de plusieurs secteurs économiques, comme on le fait en pratique statistique, on devrait retrouver une loi de distribution du <u>même type</u>.

Par exemple : la loi de distribution des chiffres d'affaires (C.A.) des épiciers devrait s'obtenir en agrégeant celles des épiciers en gros et des épiciers détaillants.
La loi de distribution des C.A. des commerces de l'alimentation devrait s'obtenir en agrégeant celles des épiciers, des bouchers, des boulangers, etc ...

Il est bien clair que le type de distribution cherché ne peut être qu'extrêment général.

La question est de savoir s'il n'existe pas de famille de distributions telle qu'en "agrégeant" deux d'entre elles (de la façon décrite au n° 1) on retrouve une distribution de la même famille.

La famille formée de toutes les distributions de variable positive possède bien cette propriété ; mais on aimerait être sûr qu'il n'existe pas des familles plus restreintes à l'intérieur de celle-ci ; c'est-à-dire des familles de distributions dépendant d'un nombre donné de paramètres.

Il faudrait trouver des fonctions $\mathcal{G}(\mathcal{P},\mathcal{Q})$ telles que l'on ait (symboliquement)

$$(5) \qquad \frac{\dfrac{1}{m}\dfrac{\partial \mathcal{G}}{\partial p}}{\dfrac{1}{(x)}\dfrac{\partial \mathcal{G}}{\partial q}} = \frac{\dfrac{1}{m_1}\dfrac{\partial \mathcal{G}}{\partial p_1}}{\dfrac{1}{(x_1)}\dfrac{\partial \mathcal{G}}{\partial q_1}} = \frac{\dfrac{1}{M}\dfrac{\partial \mathcal{G}}{\partial P}}{\dfrac{1}{(X)}\dfrac{\partial \mathcal{G}}{\partial Q}}$$

3 - ETUDE EXPERIMENTALE -

Renonçant à aborder directement un problème aussi difficile, on peut chercher des fonctions $\mathcal{G}$ de forme déterminée, en supposant qu'il existe des relations remarquables entre (x), (x_1), m, m_1.

Exemple 1 : $\mathcal{G} = \mathcal{P}^2 + Q^2 - 2Q = 0$: a-t-on la relation (5)?

(5) s'écrit :

$$\frac{\dfrac{p}{m}}{\dfrac{q-1}{(x)}} = \frac{\dfrac{p_1}{m_1}}{\dfrac{q_1-1}{(x_1)}} = \frac{\dfrac{mp + m_1 p_1}{(m + m_1)^2}}{\dfrac{(x)\,q + (x_1)\,q_1 - (x) - (x_1)}{(x)^2 + (x_1)^2}}$$

On a en réalité l'identité ci-dessous :

$$\frac{\dfrac{p}{m}}{\dfrac{q-1}{(x)}} = \frac{\dfrac{p_1}{m_1}}{\dfrac{q_1-1}{(x_1)}} = \frac{\alpha\dfrac{p}{m} + \beta\dfrac{p_1}{m_1}}{\alpha\dfrac{q-1}{(x)} + \beta\dfrac{q_1-1}{m_1}}$$

On ne parvient donc au résultat voulu que si

$$\sqrt{\alpha} = \frac{m}{m + m_1} = \frac{(x)}{(x) + (x_1)} \quad \text{(et analogue pour } \sqrt{\beta}\text{)}$$

autrement dit : $\overline{x} = \overline{x}_1$ (agrégation de groupes de même C.A. moyen).

Cette dernière relation est nécessaire et suffisante pour que l'agrégation conserve la forme de la loi de distribution.

Exemple 2 : $\mathcal{G} = (a + c)\,\mathcal{P} + (b + d)\,Q - \left[a\,\mathcal{P}^2 + (c + d)\,\mathcal{P}Q + b\,Q^2\right] = 0$

Il vient pour (2')

$$\overline{x}\,\frac{(a + c) - 2\,ap - (c + d)\,q}{(b + d) - (c + d)\,p - 2\,bq} = \overline{x}_1\,\frac{(a_1 + c_1) - 2\,a_1 p_1 - (c_1 + d_1)\,q_1}{(b_1 + d_1) - (c_1 + d_1)\,p_1 - 2b_1\,q_1}$$

En général, il est bien clair qu'on ne peut déduire (5) de (2').

Et a priori on n'a pas pensé qu'il suffisait, pour y parvenir, de faire :

$$\overline{x} = \overline{x}_1$$

Exemple 3 : $\mathcal{G} = \mathcal{P}\mathcal{Q} + \varepsilon\,\mathcal{P} - (1 + \varepsilon)\mathcal{Q} = 0$

La relation (2') s'écrit :

$$\overline{x}\,\frac{q + \varepsilon}{p - (1 + \varepsilon)} = \overline{x}_1\,\frac{q_1 + \varepsilon_1}{p_1 - (1 + \varepsilon_1)}$$

et, pour $\overline{x} = \overline{x}_1$, on peut en déduire la relation (5) ; l'agrégation semblerait donc possible, le paramètre de la distribution de l'agrégat étant

$$E = \frac{m\,\varepsilon + m_1\,\varepsilon_1}{m + m_1} = \frac{(x)\,\varepsilon + (x_1)\,\varepsilon_1}{(x) + (x_1)}$$

Ainsi $\overline{x} = \overline{x}_1$ serait une condition suffisante pour que l'agrégation conserve la loi de distribution. Il n'est pas évident que cette condition soit nécessaire; on établira pourtant dans la 2e Partie qu'elle est nécessaire et non suffisante.

DEUXIÈME PARTIE

THÉORIE DE L'AGRÉGATION DES LOIS DE DISTRIBUTION

I. - PRÉLIMINAIRES

1°) - DISTRIBUTION, LOI DE DISTRIBUTION, COURBE DE CONCENTRATION.

Il sera utile pour la suite de noter les faits suivants :

Il revient au même de se donner :

- une distribution particulière,
- la loi de distribution correspondante, plus l'effectif de la distribution,
- la courbe de concentration correspondante, plus la moyenne de la loi de distribution, plus l'effectif de la distribution.

Il revient au même de se donner :

(a) une loi de distribution particulière,
(b) la courbe de concentration correspondante, plus la moyenne de la loi de distribution.

Remarquons que ce dernier énoncé découle directement des formules déjà trouvées :

Soit x et y les coordonnées de la courbe représentant la loi de distribution,
p et q les coordonnées de la courbe de concentration,

on a :

Passage de (b) à (a)

$$\begin{cases} x = \overline{x}\dfrac{dq}{dp} \\ y = \left(\overline{x}\dfrac{d^2q}{dp^2}\right)^{-1} = f(x) \end{cases}$$

Passage de (a) à (b)

$$\begin{cases} dp = f(x)\,dx \\ dq = \dfrac{x}{\overline{x}}f(x)\,dx \end{cases}$$

Conséquence : l'agrégation concerne des distributions elles-mêmes (d'effectif m, m_1 ; de moyenne $\overline{x}, \overline{x_1}$).

Une distribution-agrégat correspond de façon univoque :

à une certaine loi de distribution,
à une certaine courbe de concentration.

Réciproquement :

une courbe de concentration donnée;
plus une moyenne donnée : $\overline{X} = (m\,\overline{x} + m_1\,\overline{x}_1)/(m + m)$,
plus un effectif donné : $M = m + m_1$,

déterminent une distribution (agrégat ou non).

2°) - NOTATIONS ET SYMBOLES :

a) On désignera par m, m_1 les effectifs ; $\overline{x}, \overline{x_1}$ les moyennes } des distributions ;

par d une distribution,
ℓ une loi de distribution,
c une courbe de distribution,
g une courbe de concentration (de Gini),

par α β .... θ des paramètres.

Aux lettres minuscules concernant les distributions de base correspondent des majuscules concernant les agrégats.

b) L'agrégat de 2 distributions de même moyenne sera symbolisé par

$$D(\bar{x}, m + m_1) = d(\bar{x}, m) + d(\bar{x}, m_1)$$

Mais on peut désigner la distribution d'effectif m, de moyenne $\bar{x}$ et admettant g comme courbe de concentration, par

$$(g, \bar{x})_m$$

Posons alors :

$$(g, \bar{x})_m + (g, \bar{x})_{m_1} = (G, \bar{X})_M$$

avec
$$M = m + m_1$$
$$\bar{X} = \bar{x}$$

autrement dit :

$$(g, \bar{x})_m + (g, \bar{x})_{m_1} = (G, \bar{x})_{m+m_1}$$

c) Le théorème de M. Fréchet (qu'on va établir plus loin) s'énonce comme suit :

1) Condition suffisante : quel que soit g, on a

$$(G, \bar{x})_{m+m_1} = (g, \bar{x})_{m+m_1}$$

2) Condition nécessaire : il n'existe aucun g tel que

$$(g, \bar{X})_{m+m_1} = (g, \bar{x})_m + (g, \bar{x}_1)_{m_1}$$

du moment que $\bar{x}$ et $\bar{x}_1$ sont distincts.

d) La distribution particulière ... $(g, \bar{x})_m$

appartient à la loi de distribution ... $(g, \bar{x}) = \ell\,(g, \bar{x})$

qui appartient elle-même à la famille de lois de distributions admettant la courbe g comme courbe de concentration, qu'on peut noter :

$$(g, \theta)$$

e) Le problème qu'on va étudier (après celui de M. Fréchet) concerne une famille de courbes de concentration $g(\alpha, \beta)$ ou $g_{\alpha\beta}$ ou g ..[à n paramètres].

On va rechercher à quelles conditions peut être réalisée une égalité de la forme :

$$(g_{AB}, \bar{X})_M = (g_{\alpha\beta}, \bar{x})_m + (g_{\alpha_1\beta_1}, \bar{x}_1)_{m_1}$$

avec
$$\begin{cases} M = m + m_1 \\ \bar{X} = \dfrac{m\,\bar{x} + m_1\,\bar{x}_1}{m + m_1} \end{cases}$$

On peut toujours poser

$$(g_{\alpha\beta}, \bar{x})_m + (g_{\alpha_1\beta}, \bar{x}_1)_{m_1} = (G, \bar{X})_M$$

G étant une courbe de concentration déterminée par les deux distributions de base, c'est-à-dire

- dépendant de la fonction g choisie,
- dépendant des valeurs α β ... et aussi α_1 β_1 ...,
- dépendant de $\overline{x}$, $\overline{x_1}$, m, m_1,

  ou plus précisément de $\frac{m}{m_1}$ et de $\frac{\overline{x}}{\overline{x_1}}$,

  ou (ce qui revient au même) de $\overline{X}$ et de M.

Le problème en vue est finalement de savoir si l'on peut trouver A B... tels que

$$G = g_{AB}$$

3°) METHODE DE CALCUL

a) Outre les notations de la 1ère Partie, on va poser :

$$\lambda = \frac{m_1}{m} \quad ; \quad \mu = \frac{(x_1)}{(x)} = \frac{m_1 \cdot \overline{x}_1}{m\overline{x}} \quad ; \qquad (1)$$

donc

$$P = \frac{p + \lambda p_1}{1 + \lambda} \quad , \quad Q = \frac{q + \mu q_1}{1 + \mu} .$$

On utilisera également :

$$k = \frac{\overline{x}}{\overline{x}_1} = \frac{\lambda}{\mu} \quad , \quad h = \frac{\overline{x}}{\overline{X}} = \frac{1 + \lambda}{1 + \mu}$$

b) Cas où l'on considère une distribution donnée :

L'idée (essentielle) prise à M. Fréchet est de choisir comme paramètre la pente de la tangente au point courant de la courbe de concentration (2), soit :

$$t = \frac{dq}{dp}$$

Les équations paramétriques de cette courbe sont donc

$$p = \varphi(t)$$
$$q = \int_a^p t\, dp = \int_a^t t\, \varphi'(t)\, dt$$

On remarquera (comme conséquence de la convexité de la courbe de concentration) que la fonction $\varphi(t)$ doit croitre de 0 à 1 lorsque t croit de a à b.

Rappelons que la limite inférieure (a) de t peut être négative mais existe (n'est pas $-\infty$) ; sa limite supérieure (b) peut ne pas exister (b : $+\infty$)

$\varphi(t)$ est encore soumis à une condition essentielle, à propos de q : Alors que, pour t = a, on a forcément q = 0, il faut s'assurer que, pour t = b, on a bien q = 1. En intégrant par parties, il vient :

$$q(t) = \Big[t\, \varphi(t)\Big]_a^t - \int_a^t \varphi(t)\, dt$$
$$q(b) = b\, \varphi(b) - \int_a^b \varphi(t)\, dt$$

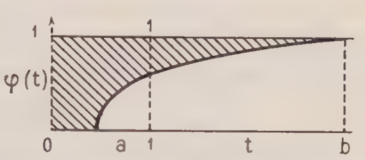
avec a $\varphi(a) = 0$, $\varphi(b) = 1$.

(1) Les effectifs m et m_1 sont en fait infiniment grands, mais λ et μ sont des nombres finis

(2) Cette courbe est donc supposée avoir des tangentes et (plus généralement) ne présenter aucune anomalie.

D'où la condition :

$$\int_a^b \varphi(t)\,dt = b - 1,$$



ce qui revient à dire que l'aire hachurée doit être égale à 1 (figure 5). Remarquons enfin qu'il revient au même de se donner la courbe de concentration g ou au contraire la fonction $\varphi(t)$ satisfaisant aux conditions précédentes.

Plus généralement, on peut remplacer la lettre g par la lettre φ dans le paragraphe 2 ci-dessus, pourvu que la fonction $\varphi(t)$ soit soumise aux conditions énoncées, qu'on appellera conditions normales (ou N) et qu'on résumera comme suit :

Conditions N pour $\varphi(t)$
$$\begin{cases} \text{croissance de } \varphi(t) \text{ de } t = a \text{ à } t = b\ ; \\ \text{limites } \varphi(a) = 0,\ \varphi(b) = 1\ ; \\ \int_a^b \varphi(t)\,dt = b - 1. \end{cases}$$

c) Cas où l'on considère 2 distributions de base et leur agrégat :

On réservera alors les lettres ($p\ q\ t\ \varphi$) à la première distribution de base ; ($p_1\ q_1\ t_1\ \varphi_1$) correspondront à la seconde et ($P\ Q\ T\ \Phi$) à l'agrégat.

Le théorème de M. Fréchet correspond au cas où $\varphi\ \varphi_1\ \Phi$ sont identiques.

La relation $x = \overline{x}\dfrac{dq}{dp}$ entraine, lorsqu'on agrège les distributions :

$$\overline{x}\frac{dq}{dp} = \overline{x}_1\frac{dq_1}{dp_1} = \overline{X}\frac{dQ}{dP}\ .$$

Mais on pose

$$\frac{dq}{dp} = t\ ,$$

donc

$$\frac{dq_1}{dp_1} = \frac{\overline{x}}{\overline{x}_1}\frac{dq}{dp}\ ,$$

ou

$$t_1 = \frac{\lambda}{\mu}t = k\,t.$$

De même

$$\frac{dQ}{dP} = \frac{\overline{x}}{\overline{X}}\frac{dq}{dp}\ ,$$

ou

$$T = \frac{1+\lambda}{1+\mu}t = h\,t\ .$$

d) Cas particulier

Si les 2 distributions de base ont même courbe de concentration, il vient :

$$\begin{cases} p_1 = \varphi(kt) \\ q_1 = \int_a^t kt\ \ \varphi(kt)\,k\,dt \end{cases}$$

Et si l'agrégat lui aussi a même courbe de concentration que les distributions de base, on a :

$$P = \varphi(ht)$$
$$Q = \int ht\ \varphi(ht)\ h\ dt$$

Il suffit de rapprocher les deux expressions de P, pour avoir la relation :

$$\varphi(ht) = \frac{p + \lambda p_1}{1 + \lambda}$$

$$\boxed{\varphi(ht) = \frac{\varphi(t) + \lambda \varphi(kt)}{1 + \lambda}} \qquad (D)$$

qui peut s'écrire :

$$\boxed{h(1-k)\ \varphi(ht) - k(1-h)\ \varphi(kt) = (h-k)\ \varphi(t)} \qquad (E)$$

Cette équation fonctionnelle (E) va jouer un rôle essentiel dans toute la théorie.

e) Cas général.

Lorsque les courbes de concentration sont distinctes, leurs équations paramétriques sont analogues, mais avec φ φ_1 et Φ au lieu de $\varphi(t)$, $\varphi(kt)$, $\varphi(ht)$.

De

$$P = \frac{p + \lambda p_1}{1 + \lambda}$$

on tire donc :

$$\boxed{\Phi(ht) = \frac{\varphi(t) + \lambda\ \varphi_1(kt)}{1 + \lambda}} \qquad (D')$$

ce qui peut s'écrire :

$$\boxed{h(1-k)\ \Phi(ht) = k(1-h)\ \varphi_1(kt) + (h-k)\ \varphi(t)} \qquad (E')$$

Cette relation définit la fonction Φ à l'aide des fonctions φ et φ_1, et des paramètres k et h.

En particulier si $\varphi(t) = \varphi(\alpha, \beta; t)$

$$\varphi_1(kt) = \varphi(\alpha_1\ \beta_1; kt)$$

la fonction Φ dépend de la fonction φ et des paramètres $(\alpha \beta \alpha_1 \beta_1 \ldots\ldots k\ h)$. Et notre problème conduit à l'équation fonctionnelle (E").

$$\boxed{h(1-k)\ \varphi(A, B; ht) - k(1-h)\ \varphi(\alpha_1\ \beta_1\ ; kt) = (h-k)\ \varphi(\alpha\beta\ ;\ t)} \qquad (E'')$$

qui fait pendant à la formule D"

$$\varphi(A, B; ht) = \frac{\varphi(\alpha, \beta\ ; t) + \lambda\ \varphi(\alpha_1 \beta_1\ ; kt)}{1 + \lambda} \qquad (D'')$$

f) Conditions supplémentaires ou conditions S :

On désignera par "conditions S" imposées à $\varphi(t)$, les conditions : $a = 0$, $b = +\infty$, autrement dit (fig.6)

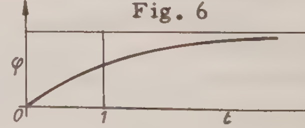


Fig. 6

$$\varphi(0) = 0 \qquad \varphi(+\infty) = 1$$

Ces conditions jouent un rôle essentiel.

D'ores et déjà, on voit que, puisque

$$t = \frac{dq}{dp} = \frac{x}{\overline{x}}$$

les "conditions S" signifient que la variable x a comme intervalle de variation $(0, +\infty)$.

II. - CAS DE DISTRIBUTIONS ADMETTANT LA MÊME COURBE DE CONCENTRATION

1°) - RESUME DES CALCULS QUI SUIVENT

Le problème qui se pose est la résolution de l'équation fonctionnelle E;

$$h(1-k)\,\varphi(ht) - k(1-h)\,\varphi(kt) = (h-k)\,\varphi(t) \qquad (E)$$

où $\varphi(t)$ doit déjà remplir les conditions N.

On va d'abord supposer h et k différents de 1 ; et on va montrer qu'aucune fonction $\varphi(t)$ n'existe; on sera ramené alors au cas où

$$k = 1$$

ce qui entraine $k = h$; et alors (E) est manifestement satisfaite quelque soit $\varphi(t)$. De même que (D).

Tel est le théorème de M. Fréchet.

Supposons donc $k \neq h \neq 1$.

2°) - LES CONDITIONS S SONT EN REALITE DES CONDITIONS NECESSAIRES -

On voit facilement qu'aucune fonction φ n'existe si t n'est pas une variable prenant toutes les valeurs de 0 à $+\infty$.

En effet, si $0 \leqslant a \leqslant t \leqslant b$, ceci veut dire que x varie entre

$a\overline{x}$ et $b\overline{x}$ pour la 1ère distribution

$a\overline{x}_1$ et $b\overline{x}_1$ pour la 2e distribution

$a\overline{X}$ et $b\overline{X}$ pour leur agrégat.

Supposons $\overline{x}_1 < \overline{x}$ (donc $k > 1$ et $h > 1$)

donc $a\overline{x}_1 < a\overline{X} < a\overline{x}$ et $b\overline{x}_1 < b\overline{X} < b\overline{x}$.

Il existerait ainsi des unités de la 2e population pour lesquelles x prendrait une valeur comprise entre $a\overline{x}_1$ et $a\overline{X}$; alors que, dans l'agrégat des deux populations, il n'existerait plus de telles unités. Ceci oblige à prendre $a = 0$.

De même les unités de la 1ère population, pour lesquelles x serait compris entre $b\overline{X}$ et $b\overline{x}$, disparaitraient quand on envisage l'agrégat des deux populations. Ceci ne peut s'expliquer que si b est infini.

Ainsi $\varphi(t)$ devra remplir à la fois les conditions N et S.

3°) - RESOLUTION DE (E)

a) Première méthode.

On remarque que $\qquad \varphi(t) = 1$

et $\qquad \varphi(t) = \frac{1}{t}$

sont 2 solutions particulières de l'équation (E); mais celle-ci est "linéaire" de sorte que

$$\varphi(t) = \varphi_0(t) = \lambda + \frac{\mu}{t}$$

(où λ et μ sont 2 paramètres arbitraires) est aussi une solution de (E); mais (E) est une équation aux différences finies du 2e ordre, la variable étant $\mathcal{L}t$. Ainsi $\varphi_0(t)$ est l'intégrale générale de (E).

b) Seconde méthode.

En supposant que les fonctions $\varphi(t)$ sont dérivables, on peut résoudre directement l'équation (E), ce qui nous servira par la suite.

Dérivons les 2 membres de (E) par rapport à h, puis par rapport à k. Il vient :

$$(1-k)\varphi(ht) + h(1-k)t\varphi'(ht) + k\varphi(kt) = \varphi(t)$$

$$-\varphi(ht) - ht\varphi'(ht) + \varphi(kt) + kt\varphi'(kt) = 0$$

C'est-à-dire :

$$\varphi(ht) + ht\varphi'(ht) = \varphi(kt) + kt\varphi'(kt).$$

Et, comme ($k \neq h$) par hypothèse, il vient

$$\varphi(ht) + ht\varphi'(ht) = \text{constante}$$

équation de la forme

$$\varphi(u) + u\varphi'(u) = \lambda$$

dont l'intégrale générale est

$$\varphi(u) = \lambda + \frac{\mu}{u}$$

4°) - CONCLUSION

Avec $k \neq 1$, il n'existe d'autres fonctions dérivables $\varphi(t)$, solutions de (E), que celles de la forme

$$\lambda + \frac{\mu}{t} = \lambda - \frac{|\mu|}{t}$$

croissantes pour $\mu < 0$, mais qu'on vérifie être incompatibles avec les conditions (S); car φ tend vers $-\infty$ quand t tend vers 0.

On est donc ramené au cas où $k = 1 = h$, c'est-à-dire $\overline{x} = \overline{x}_1 = \overline{X}$; alors tout φ convient.

Ainsi la condition nécessaire et suffisante pour avoir

$$(g, \overline{X})_{m+m_1} = (g, \overline{x})_m + (g, \overline{x_1})_{m_1}$$

est que $$\overline{x} = \overline{x}_1 \; (= \overline{X})$$

Ce théorème de M. Fréchet s'étend tout de suite à un nombre quelconque de distributions de base, possédant toutes la même courbe de concentration.

5°) - PORTEE DU RESULTAT -

a) On aurait pu imaginer que la courbe G pouvait (pour certaines courbes g) venir coincider avec g pour quelque combinaison privilégiée de $\overline{x}\,\overline{x}_1$ m et m_1. Ainsi il n'en est rien.

b) On aurait pu penser qu'il existait peut-être des courbes de concentration g, telles que les courbes de distribution C_θ des lois de distributions

$$(g, \theta)$$

formaient un faisceau, c'est-à-dire se coupaient en un nombre fini de points fixes. Car il est clair que la courbe de distribution de

$$(G, \overline{X})$$

passe par les points de rencontre des courbes de distribution de

$$(g, \overline{x}) \text{ et } (g, \overline{x}_1)$$

et (lorsque m et m_1 varient) forme un véritable faisceau.

Autrement dit, l'équation de C_θ étant de la forme

$$y - \frac{1}{\theta}\, f\left(\frac{x}{\theta}\right) = 0$$

on pouvait se demander s'il n'était jamais possible d'écrire (avec une forme adéquate de fonction f)

$$\frac{1}{\overline{x}}\, f\left(\frac{x}{\overline{x}}\right) - \frac{1}{\overline{x}_1}\, f\left(\frac{x}{\overline{x}_1}\right) \equiv \xi(x) \, . \, \eta\, (\overline{x}, \overline{x}_1)$$

On vient donc d'établir que cela n'était jamais possible.

Remarque : Dans la le Partie, la condition nécessaire et suffisante $\overline{x} = \overline{x}_1$ avait bien été trouvée avec la courbe particulière g, d'équation :

$$p^2 + q^2 - 2\,q = 0.$$

III. - CAS D'UNE FAMILLE DE COURBES DE CONCENTRATION

1°) - FAMILLES G CONSTRUITES A PARTIR DE g ET g_1 :

a) Partons de 2 courbes de concentration g et g_1 distinctes et fixes, d'ailleurs quelconques. La relation (E')

$$h(1-k)\,\Phi(ht) = k(1-h)\,\varphi_1(kt) + (h-k)\,\varphi(t) \qquad (E')$$

définit Φ, donc G. Lorsque h et k varient, on engendre ainsi toute une famille de courbes de concentration G à 2 paramètres (h,k).

On voit que si $h \longrightarrow 1$, (E') donne à la limite

$$\Phi(t) = \varphi(t)$$

tandis que si $h \longrightarrow k$, (E') donne

$$\Phi(kt) = \varphi_1(kt)$$

Dans le 1er cas : G tend vers g, et dans le second G tend vers g_1 : de sorte que la famille G_{hk} renferme bien g et g_1.

b) Par exemple : laissons fixes $\overline{x}$ et $\overline{x}_1$, c'est-à-dire k, mais faisons varier m et m_1

C'est-à-dire $$\lambda = \frac{m_1}{m}$$

ou $$h = \frac{1+\lambda}{1+\dfrac{\lambda}{k}}$$

On a une famille G_λ à 1 paramètre qui renferme g et g_1.

c) Lien avec le problème étudié ici :

Malheureusement la famille G_{kh} ne constitue pas nécessairement une solution au problème posé ici, lequel est de trouver une famille $g_{\lambda\mu}$ com-

prenant ses propres agrégats G. On a seulement déplacé le problème; car il faudrait à présent se demander si l'agrégat de G_{kh} et $G_{k'h'}$ appartient ou non à la famille

$$G_{\alpha\beta}$$

C'est pourquoi on n'abordera pas ainsi la question ; on supposera au contraire le problème résolu, c'est-à-dire qu'on admettra l'existence d'une famille de courbes $g_{\alpha\beta}$, autrement dit de fonctions

$$\varphi(\alpha, \beta ; t)$$

satisfaisant à l'équation fonctionnelle (E'') et aux conditions annexes avec $k \neq 1$. Et on essayera de trouver l'expression de φ, ce qui conduira à établir que φ ne peut exister. On retombera alors sur $k = 1$, d'où $k = h$ et (E'') vérifiée quel que soit φ

Mais (à la différence du II ci-dessus), le fait que (E'') soit vérifiée n'implique pas du tout que D'' le soit (MM. Fréchet et Malinvaud nous ont donné des exemples simples où (E'') n'entraine pas l'agrégation, c'est aussi le cas pour nos exemples n° 2 et 3). On en conclura que :

a) $h=k=1$ (c'est-à-dire $\bar{x} = \bar{x}_1$) est une condition nécessaire à l'agrégation;

b) (D'') donne la forme générale des courbes de concentration qui se prêtent à l'agrégation.

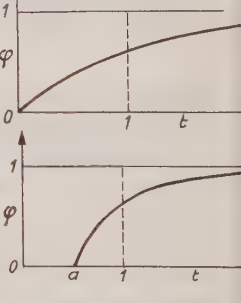
2°) - CONDITIONS IMPOSEES AUX SOLUTIONS φ DE L'EQUATION FONCTIONNELLE (E'')

$$h(1-k)\,\varphi\,(A,B ; ht) - k\,(1-h)\,\varphi\,(\alpha_1, \beta_1 ; kt) = (h-k)\,\varphi\,(\alpha, \beta ; t) \quad (E'')$$

Conditions N : sans changement.

La fonction φ doit être croissante, de $\varphi(a) = 0$ à $\varphi(b) = 1$. En outre on doit avoir

$$b - \int_a^b \varphi(t)\,dt = 1$$



Conditions S :

L'intervalle de variation de x doit être le même pour les lois de distribution $(g, \bar{x})$, $(g_1, \bar{x}_1)$ et $(G, \bar{X})$; d'où les conditions:

$$(\min.\ x) = a\,\bar{x} = a_1\,\bar{x}_1 = A\,\bar{X}$$

$$(\max.\ x) = b\,\bar{x} = b_1\,\bar{x}_1 = B\,\bar{X}$$

D'où 2 grands cas possibles :

1) Ou bien $\bar{x}\ \bar{x}_1\ \bar{X}$ sont arbitraires, mais on a $\begin{cases} a = 0 \\ b + \infty \end{cases}$

2) Ou bien a n'est pas nul et b est fini; mais $\bar{x}_1$ et $\bar{X}$ sont liés à $\bar{x}$.

Ainsi, à côté du 1er cas, analogue à celui étudié au II, s'ouvre un 2ème cas où se restreignent les exigences du problème.

3°) - RESOLUTION DE (E'') - PREMIERE METHODE -

Comme pour (E), il est clair que $\varphi = \lambda$ et $\varphi = \mu/t$ sont des solutions particulières de (E''), - à condition cette fois que λ et μ soient des fonctions convenables de α, β :

C'est-à-dire si on a :

$$h(1-k)\,\lambda\,(A,B) = k(1-h)\,\lambda\,(\alpha_1, \beta_1) + (h-k)\,\lambda\,(\alpha, \beta) \ldots \quad (R_1)$$

et aussi :

$$(1-k)\,\mu\,(A,B) = (1-h)\,\mu\,(\alpha_1, \beta_1) + (h-k)\,\mu\,(\alpha, \beta) \ldots \quad (R_2)$$

pour des expressions A, B, convenables de α, β, α_1, β_1, k et h. Alors

$$\varphi(t) = \lambda + \frac{\mu}{t} = \lambda - \frac{|\mu|}{t} \quad (\text{si } \mu < 0)$$

est la solution générale (décroissante) de (E'').

<u>1er cas</u>. Cette solution ne convient évidemment pas dans le 1er cas (où les conditions S sont exigées); car, quels que soient par ailleurs λ et μ, $\varphi(t)$ tend vers $-\infty$ (et non pas vers 0) quand t tend vers zéro.

Il en résulte qu'il n'existe aucune solution pour $k \neq 1$.

En revanche, lorsque k = 1, c'est-à-dire $\overline{x} = \overline{x}_1$, on a

$$h = 1$$

et l'équation (E'') est identiquement vérifiée, sans entrainer pourtant (D''):

Il n'y a aucune raison, en effet, pour qu'on ait identiquement :

$$\varphi(A, B; t) = \frac{\varphi(\alpha, \beta; t) + \lambda\varphi(\alpha_1, \beta_1; t)}{1 + \lambda}, \quad (\text{avec } \lambda = \frac{m_1}{m})$$

pour un système convenable de paramètres A, B (lorsque α β α_1 β_1 sont donnés).

On retrouve donc le <u>théorème</u> de M. Fréchet, mais amputé : la condition nécessaire pour avoir :

$$(g_{AB}, \overline{X})_{m+m_1} = (g_{\alpha\beta}, \overline{x})_m + (g_{\alpha_1\beta_1}, \overline{x}_1)_{m_1}$$

<u>lorsque la variable x prend toutes valeurs positives</u>, est que

$$\overline{x} = \overline{x}_1$$

(résultat qui s'étend à un agrégat de n distributions).

<u>2° cas</u>. En revanche la solution de (E'')

$$\varphi(t) = \lambda - \frac{|\mu|}{t}$$

parait pouvoir convenir dans le 2ème cas. Il vient :

$$\left.\begin{array}{ll} t = a \; ; & \lambda - \frac{|\mu|}{a} = 0 \\ t = b \; ; & \lambda - \frac{|\mu|}{b} = 1 \end{array}\right\} \quad \begin{array}{l} |\mu| = \frac{ab}{b-a} \\ \lambda = \frac{b}{b-a} \end{array}$$

d'où :

$$\varphi(t) = \frac{b}{b-a}\left(1 - \frac{a}{t}\right)$$

On peut considérer que $a = \alpha$ est <u>le paramètre</u> de la courbe g.

La relation entre les distributions associées $(g, \overline{x})_m$ et $(g_1, \overline{x}_1)_{m_1}$ est

$$\alpha\overline{x} = \alpha_1\overline{x}_1$$

donc $\alpha_1 = k\alpha$ ou $\overline{x}_1 = \overline{x}\dfrac{\alpha}{\alpha_1}$.

Pour avoir la relation analogue avec la limite inférieure b, il faut que

$$b = r\alpha$$
$$b_1 = r\alpha_1$$

r étant une <u>constante</u>, <u>indépendante</u> de la courbe g choisie.

L'expression $\varphi(t)$ s'écrit alors

$$\varphi(t) = \frac{t}{r-1}\left(1 - \frac{\alpha}{t}\right)$$

Assurons nous que (R_1) et (R_2) sont bien vérifiées : On a

$\lambda(\alpha,\beta) = \frac{r}{r-1}$ (constante absolue); donc (R_1) est vérifiée

$\mu(\alpha,\beta) = -\frac{r\alpha}{r-1}$ (R_2) fixe la valeur du paramètre de G lorsque α et α_1 sont donnés.

$$(1-k)\,A = (1-h)\,\alpha_1 + (h-k)\,\alpha$$

Mais $\frac{\overline{\overline{x}}}{\overline{x}} = k = \frac{\alpha_1}{\alpha}$ d'où $\underline{A = \alpha h}$

Reste la condition N des aires :

$$1 = b - \int_a^b \varphi(t)\,dt,$$

C'est-à-dire :

$$1 = r\alpha - \frac{r}{r-1}\int_\alpha^{r\alpha}\left(1 - \frac{\alpha}{t}\right)dt,$$

d'où :

$$\alpha = \frac{r-1}{r\,\mathcal{L}\,r}$$

Conclusions : Nous obtenons une distribution unique et non pas une famille G_α, puisque r doit rester le même (afin d'assurer une limite supérieure commune de la variable x à toutes les distributions) tandis que α doit varier.

Il est impossible d'ailleurs de faire tendre r vers l'infini, comme on le voit bien sur les équations

de la courbe de concentration
$$\begin{cases} p = \varphi(t) = \frac{r}{r-1} - \frac{t}{t\,\mathcal{L}\,r} \\ q = \frac{1}{\mathcal{L}r}\,\mathcal{L}(-) = \frac{1}{\mathcal{L}r}\,\mathcal{L}\left[\frac{r\,\mathcal{L}\,r}{r-1}\,t\right] \end{cases}$$

ou de la courbe de fréquences
$$\begin{cases} x = \overline{x}\,t \\ y = 1/\overline{x}\,t^2\mathcal{L}\,r \end{cases}$$

d'où

$$y = \left(\frac{\overline{x}}{\mathcal{L}r}\right)\frac{1}{x^2}$$

Il s'agit d'une loi du type de Pareto, mais tronquée car l'exposant de x (x^{-2}) est trop petit pour assurer la convergence de l'intégrale

$$\int_\alpha^{+\infty} x\,y\,dx$$

Le 2ème cas n'aboutit donc pas plus que le 1er cas (en dehors du cas $h = k = 1$).

4°) - RESOLUTION DE (E''). - SECONDE METHODE -

Partons encore de (E'') où par hypothèse on a $k \neq 1$.

$$h(1-k)\,\varphi(A,B\,;\,ht) = k(1-h)\,\varphi(\alpha_1,\beta_1\,;\,kt) + (h-k)\,\varphi(\alpha,\beta\,;\,t) \qquad (E'')$$

avec
$$A = A(\alpha,\beta\,;\,\alpha_1\,\beta_1\,;\,h,k)$$
$$B = B(\alpha,\beta\,;\,\alpha_1\,\beta_1\,;\,h,k)$$

La relation (E'') est supposée vérifiée pour toute valeur de t, lorsque $(\alpha\ \beta\ \alpha_1 \beta_1\ ; h\ k)$ sont donnés par ailleurs.

Retranchons terme à terme de (E'') l'expression suivante :

$$h\,(1-k)\ \varphi\ (\alpha, \beta\ ; t) \equiv k(1-h)\ \varphi\ (\alpha, \beta\ ; t) + (h-k)\ \varphi\ (\alpha, \beta\ ; t)$$

Il vient :

$$h(1-k)\left[\varphi(A,B; ht) - \varphi(\alpha, \beta\ ; t)\right] = k(1-h)\left[\varphi(\alpha_1 \beta_1\ ; kt) - \varphi(\alpha,\beta; t)\right]$$

Cette identité en t n'est possible que si :

$$\varphi(A,B\ ; ht) - \varphi(\alpha,\beta;t) = \phi(t)\ \theta(A,B\ ; \alpha, \beta\ ; h)$$

$$\varphi(\alpha_1, \beta_1, kt) - \varphi(\alpha,\beta;t) = \phi(t)\ \theta(\alpha_1\beta_1\ ; \alpha,\beta\ ; h)$$

avec

$$h(k-1)\ \theta\ (A,B\ ; \alpha, \beta\ ; h) = k(h-1)\,\theta\ (\alpha_1, \beta_1 ; \alpha, \beta\ ; k)$$

c'est-à-dire :

$$\frac{h}{h-1}\ \theta\ (A,B;\alpha, \beta\ ; h) = \frac{k}{k-1}\ \theta\ (\alpha_1\ \beta_1 ; \alpha, \beta;k)$$

Cette expression doit donc être une constante "<u>absolue</u>" $\lambda(\alpha, \beta)$ quelle que soit la distribution variable considérée (d_1, ou D); ainsi on doit avoir

$$\theta\ (\alpha_1, \beta_1, \alpha, \beta\ ; k) = \left(\frac{k-1}{k}\right) \lambda\ (\alpha, \beta\ )$$

d'où

$$\boxed{\varphi(\alpha_1, \beta_1 ; kt) - \varphi\ (\alpha, \beta, t) = \frac{k-1}{k}\ \lambda(\alpha,\beta)\,\phi\ (t)} \qquad \text{(H)}$$

On est ramené à une distribution à <u>un seul paramètre</u>, <u>à savoir</u> $k = \frac{\overline{x}}{\overline{x}_1}$ (où $\overline{x}_1$ varie et $\overline{x}$ reste constant); car le paramètre λ lié à la courbe $g(\alpha\ \beta)$ est déterminé par <u>la condition N des aires</u> ; si $a \leqslant t \leqslant b$, on a

$$\int_a^b \varphi(t)\ dt = b - 1$$

$$\int_{ka}^{kb} \varphi(kt)\ k\ dt = kb - 1$$

En intégrant les 2 membres de (H) entre a et b, il vient :

$$\frac{kb-1}{k} - (b-1) = \frac{k-1}{k} \cdot \lambda \cdot \left[\Psi\ (b) - \Psi\ (a)\right]$$

où l'on a posé $$\Psi\ (t) = \int \phi\ (t)$$

On peut mettre en facteur $(1 - 1/k)$; et il reste simplement l'expression de λ :

$$\lambda^{-1} = \Psi\ (b) - \Psi\ (a)$$

expression qui ne dépend donc plus que des bornes (a,b) de l'intervalle de variation de t.

Ainsi lorsqu'on possède une solution $\varphi(t)$ de l'équation (E''), on a [Voir note (1)]:

$$\varphi(kt) = \varphi(t) + \frac{k-1}{k} \frac{\Psi'(t)}{\Psi(b) - \Psi(a)}$$

Dérivons par rapport à k ; il vient :

$$t\,\varphi'(kt) = \frac{1}{k^2} \frac{\Psi'(t)}{\Psi(b) - \Psi(a)} \quad ,$$

d'où l'équation fonctionnelle en φ

$$\varphi(kt) - \varphi(t) = k(k-1)\, t\, \varphi'(kt)$$

Dérivons une seconde fois par rapport à k ; il vient :

$$t\,\varphi'(kt) = (2k-1)\, t\, \varphi'(kt) + k(k-1)\, t^2\, \varphi''(kt)$$

d'où, si $(k-1) \neq 0$, et en simplifiant :

$$2\,\varphi'(kt) + kt\,\varphi''(kt) = 0$$

c'est-à-dire

$$2\,\varphi'(u) + u\,\varphi''(u) = 0$$

équation dont l'intégrale générale est

$$\varphi(u) = \lambda + \frac{\mu}{u}$$

Conditions S :

On a vu (au II) qu'avec $a = 0$ et $b\ \infty$, cette solution ne convenait pas au problème (1er cas).

Mais on n'a aucune raison d'imposer cette condition; en adoptant (2e cas)

$$a = \alpha \quad , \quad b = r\alpha$$

on retrouvera la distribution donnée au paragraphe 3; laquelle est à rejeter puisqu'il s'agit d'une distribution isolée et non pas d'une famille de distributions.

Alors $\Psi(b) - \Psi(a) = \mathcal{L}(r\alpha) - \mathcal{L}\alpha = \mathcal{L}r$, ne dépend plus de α.

Seul subsiste donc le cas où $k = 1$, c'est-à-dire $\overline{x} = \overline{x}_1$, où (E'') est vérifiée d'emblée, sans que cela implique que (D'') le soit.

Concluons donc que seule la condition nécessaire de M. Fréchet s'étend, sans aucune restriction, aux familles de distribution à un nombre quelconque de paramètres.

---

(1) Note : En particulier, avec

$$\varphi(t) = \frac{r}{r-1}\left(1 - \frac{\alpha}{t}\right) = \frac{r}{r-1} - \frac{1}{Lr.t}$$

on a

$$\varphi(kt) = \frac{r}{r-1} - \frac{1}{Lr.kt}$$

donc

$$\varphi(kt) = \varphi(t) + \frac{1}{Lr.t}\left(\frac{-1}{k} + 1\right)$$

$$= \varphi(t) + \frac{k-1}{k} \frac{(Lt)'}{L(\alpha r) - L\alpha}$$

5°)- REMARQUE :

La possibilité de courbes de concentration confondues avec les segments

$$0 \leqslant p < 1 \qquad q = 0$$
$$p = 1 \qquad 0 < q \leqslant 1$$

n'a pas été non plus envisagée ; c'est un cas limite <u>acceptable</u> de $\varphi(t) = \lambda - \frac{|\mu|}{t}$

6°)- THEOREME :

S'il existe une famille de courbes de concentration qui se prêtent à l'agrégation, elles sont nécessairement tangentes entre elles aux deux extrémités de leur arc ($p = q = 0$; $p = q = 1$).

<u>Démonstration</u> : On a établi au n° 2 qu'on avait

$$a\overline{x} = a_1\,\overline{x}_1 = A\overline{X} \qquad b\overline{x} = b_1\,\overline{x}_1 = B\overline{X}$$

et on vient de démontrer que $\overline{x} = \overline{x}_1 = \overline{X}$; d'où il suit que

$$a = a_1 = A \qquad b = b_1 = B$$

<u>Une conséquence</u> : la famille de distributions de l'exemple 3 (voir à la fin de la 1ère partie ci-dessus) ne se prête pas à l'agrégation, puisque a et b dépendent de α .

Effectivement on a : $pq + \varepsilon\, p\,(1+\varepsilon)\, q = 0$

et en dérivant : $pq' + q + \varepsilon - (1+\varepsilon)\, q' = 0$

d'où pour $\begin{cases} p = q = 0 & a = q' = \varepsilon/(1+\varepsilon) \\ p = q = 1 & b = q' = (1+\varepsilon)/\varepsilon \end{cases}$

7°)- CONSTRUCTION DES FAMILLES DE DISTRIBUTIONS QUI SE PRETENT A L'AGREGATION - Familles à 1 paramètre -

Reprenons le procédé du n° 1ab ci-dessus, avec 2 <u>courbes fixes</u> de concentration g et g_1 ayant un contact aux 2 extrémités (d'après le 6°), en supposant $\overline{x}$ et $\overline{x}_1$ fixes et en outre égaux (d'après les 3° et 4°), et en faisant varier le rapport des effectifs :

$$\lambda = m_1/m$$

La relation (D'') détermine une fonction Φ de t, variant de 0 à 1 :

$$P = \Phi = \frac{\varphi + \lambda\varphi_1}{1 + \lambda} \qquad \text{(D'')}$$

et il est immédiat que, sur l'autre axe des coordonnées, on a de même

$$Q = \Psi = \frac{\phi + \lambda\phi_1}{1 + \lambda}$$

(λ au lieu du μ figurant dans Q en I-3°). Autrement dit, à chaque λ correspond une courbe de concentration G_λ et une seule, définie géométriquement comme suit :

a) On mène la tangente à g et celle à g_1 de pente t ;

b) On joint les 2 points de contact c et c_1 ;

c) On prend sur la droite cc_1, le point C tel que $\frac{\overline{cC}}{\overline{cc_1}} = \frac{\lambda}{1+\lambda}$;

d) On fait varier t (de a jusqu'à b) de façon que c décrive g, c_1 décrive g_1 et C décrive G_λ.

En particulier G_0 est g et G_0 est g_1.

Il reste cependant à démontrer que <u>la famille G_λ se prête bien à l'agrégation</u>.

Ceci résulte du seul fait que, quelque soit λ, la tangente à G_λ en C <u>a la même pente t</u> (indépendante de λ), comme on va l'établir.

En effet $\vec{c}_1 = \vec{c} + \vec{u}\rho$;

donc $d\vec{c}_1 = d\vec{c} + d(\vec{u}\rho)$ avec $d\vec{c}_1 = d\vec{c}$,

c'est-à-dire $d(\vec{u}\rho) = 0$.

Mais $\vec{C} = \vec{c} + \vec{u}\rho\,(\lambda/1+\lambda)$,

donc $d\vec{C} = d\vec{c} + d(\vec{u}\rho)\,(\lambda/1+\lambda)$,

donc $d\vec{C} = d\vec{c}$.

Ce point établi, il est immédiat que l'agrégation portant sur G_{λ_0} et G_{λ_1} avec des effectifs M_0 et M_1 redonnera une courbe de la famille G (en supposant toujours les moyennes identiques).

<u>Conclusion</u> : Etant donné deux courbes de concentration <u>arbitraires</u> g et g_1 mais en contact aux extrémités, la famille G_λ définie par les conditions

$$\overrightarrow{cC} = \overrightarrow{cc_1}\,(\lambda/1+\lambda) \text{ avec } d\vec{C} = d\vec{c},$$

se prête à l'<u>agrégation à moyennes égales</u>.

<u>Réciproquement</u> s'il existe une famille de distributions se prêtant à l'agrégation :

les moyennes des distributions sont égales entre elles,

les courbes de concentration ont un contact entre elles aux 2 extrémités,

toute courbe de concentration de la famille peut se déduire de 2 certaines d'entre elles par le procédé décrit précédemment.

<u>Exemple</u> : Avec notre exemple 3, on avait :

$$q = (1+\varepsilon)\,t - \varepsilon - pt,$$

$$q(p-1-\varepsilon) + \varepsilon p = 0,$$

d'où
$$p = 1 + \varepsilon - \sqrt{\frac{\varepsilon(1+\varepsilon)}{t}} = \varphi_1(t) ;$$

posons
$$1 + \varepsilon_1 - \sqrt{\frac{\varepsilon_1(1+\varepsilon_1)}{t}} = \varphi_1(t),$$

$$E = \frac{\varepsilon + \lambda\varepsilon_1}{1+\lambda} ;$$

il vient
$$1 + E + \sqrt{\frac{A}{t}} = \frac{\varphi + \lambda\varphi_1}{1+\lambda},$$

avec
$$\frac{\sqrt{\varepsilon(1+\varepsilon)} + \lambda\sqrt{\varepsilon_1(1+\varepsilon_1)}}{1+\lambda} = \sqrt{A}.$$

Il n'y a aucune raison pour que cette dernière expression soit identique à

$$\frac{\sqrt{(\varepsilon+\lambda\varepsilon_1)\left[(1+\varepsilon)+\lambda(1+\varepsilon_1)\right]}}{1+\lambda} = \sqrt{E(1+E)}.$$

<u>Autre exemple</u> (dû à M. MALINVAUD) : Avec une loi de Pareto, on a :

$$p = 1 - \left[\alpha/(\alpha-1)\,t\right]^{\alpha} = \varphi(t)$$

Si l'on prend une 2ème loi, de paramètre il est clair qu'on a

$$\frac{\varphi + \lambda\varphi_1}{1+\lambda} = 1 - \left(\frac{B}{t^{\alpha}} + \frac{C}{t^{\alpha_1}}\right)$$

et qu'on ne peut trouver A tel que la parenthèse soit de la forme

$$\frac{D}{t^{A}}$$

8°)- PROBLEME

Reconnaitre si une famille de courbes de concentration (à 1 paramètre α) correspond à des distributions qu'on peut agréger (sous réserve que les moyennes soient égales).

<u>Solution</u> : On éliminera le paramètre α entre les équations

$$g(\alpha\ ;\ p,q) = 0$$

$$g'_p(\alpha\ ;\ p,q) + t\, g'_q(\alpha\ ;\ p,q) = 0$$

Le résultat de cette élimination doit représenter (entre autres) une droite (pour chaque valeur de $t = dq/dp$).

<u>Exemple</u> : Faisceau de coniques bitangentes en ($p=q=0$), ($p=q=1$).

$$g(p,q) = q(1-p) + \alpha\ (q-p)^2 = 0$$

Ces coniques ont un arc ($0 \leqslant p \leqslant 1,\ 0 \leqslant q \leqslant 1$) qui satisfait à la condition nécessaire de contact aux extrémités. Les distributions correspondantes seraient-elles susceptibles de s'agréger ?

La 2ème équation s'écrit :

$$\left[t(1-p)-q\right] + 2\alpha(q-p)(t-1) = 0$$

L'élimination de α donne :

(1) $$q - p = 0$$

équation d'une droite qui ne convient pas :

(2) $$2(t-1)\,q(1-p) - (q-p)\left[t(1-p)-q\right] = 0\ ,$$

équation d'une conique qui se décompose si elle a un point double :

$$\left.\begin{array}{ll} (A) & -2(t-1)\,q + t\,(1-p) - q + (q-p)t = 0 \\ (B) & 2(t-1)\,(1-p) - t\,(1-p) + q + q - p = 0 \\ (C) & 2(t-1)\,q \qquad -t(q-p) = 0 \end{array}\right\}\ \text{Equations du point double}$$

(A) et (C) donnent (U) $$t(1-p)-q = 0$$

(B) et (U) donnent (V) $$2(t-1)(1-p) + (q-p) = 0$$

$$\left.\begin{array}{ll} \text{(V) et (C) donnent} & \dfrac{q-p}{2(t-1)} = \dfrac{q}{t} = p-1 \\ \text{alors que (U) donne} & \dfrac{q}{t} = 1 - p \end{array}\right\}\ p = 1,\ q = 0,\ p = q$$

<u>Conclusion</u> : Il n'y a pas point double ; le faisceau de coniques ne représente donc pas des distributions qui peuvent s'agréger.

9°) CONSTRUCTION DES FAMILLES DE DISTRIBUTION QUI SE PRETENT A L'AGREGATION - Familles à plusieurs paramètres.

Le procédé n° 7 s'étend à un nombre quelconque de paramètres. Par exemple, si g_2 est une 3° courbe de concentration ayant un contact en ses extrémités avec g et g_1 et n'appartenant pas à la famille G_λ, on peut construire une famille de courbes de concentration $G_{\lambda\mu}$ à 2 paramètres en posant maintenant

$$\left\{\begin{array}{l} \mathcal{P} = \dfrac{P + \mu\varphi_2}{1+\mu} \\ \mathcal{Q} = \dfrac{Q + \mu\phi_2}{1+\mu} \end{array}\right. \quad \text{avec} \quad \left\{\begin{array}{l} P = \dfrac{\varphi + \lambda\varphi_1}{1+\lambda} \\ Q = \dfrac{\phi + \lambda\phi_1}{1+\lambda} \end{array}\right. ;$$

c'est-à-dire :

$$\left\{\begin{array}{l} \mathcal{P} = \alpha\varphi + \beta\varphi_1 + \gamma\varphi_2 \\ \mathcal{Q} = \alpha\phi + \beta\phi_1 + \gamma\phi_2 \end{array}\right. \qquad \left\{\alpha + \beta + \gamma = 1,\right.$$

avec

$$\alpha = \frac{1}{(1+\lambda)(1+\mu)}, \quad \beta = \frac{\lambda}{(1+\lambda)(1+\mu)}, \quad \gamma = \frac{\mu}{1+\mu}.$$

On démontre de même que :

L'agrégat des distributions $(G_{\lambda\mu}, \bar{x})_M$ et $(G_{\lambda'\mu'}, \bar{x})_{M'}$ est de la forme $(G_{\lambda''\mu''}, \bar{x})_{M+M'}$. Il est plus commode d'employer les 3 paramètres $\alpha\beta\gamma$ que les 2 paramètres $\lambda\mu$; en posant

$M'/M = \rho$, les 3 paramètres de l'agrégat sont $\left(\dfrac{\alpha+\rho\alpha'}{1+\rho}, \dfrac{\beta+\rho\beta'}{1+\rho}, \dfrac{\gamma+\rho\gamma}{1+\rho}\right)$

de somme 1.

10°) RETOUR A LA LOI DE DISTRIBUTION : $p = \varphi(t) = \varphi\left(\frac{x}{\bar{x}}\right)$

Etant donné les lois de distribution $p(x)$, $p_1(x)$, $p_2(x)$, on retrouve un résultat très banal :

$$\mathcal{P} = \alpha p + \beta p_1 + \gamma p_2 \qquad (\text{avec } \alpha + \beta + \gamma = 1)$$

On retrouve en même temps un <u>théorème</u>, essentiel pour d'autres problèmes d'agrégation :

Pour qu'on ait une identité de la forme

$$F(x,r) + \lambda F(x,r_1) \equiv (1+\lambda)F\left[x, R(r,r_1)\right]$$

il faut et il suffit qu'on ait

$$F(x,r) \equiv F_1(x) + G(r)F_2(x)$$

Ce théorème s'établit en dérivant l'identité par rapport au paramètre r et en constatant que ($\log F'_r - \log F''_r$) est indépendant de x. C'est à M. BAUDIER que nous devons de faire ce rapprochement entre les 2 théories et de pouvoir énoncer finalement que :

Se prêtent seules à l'agrégation les combinaisons <u>linéaires</u> de distributions ayant <u>même moyenne</u> et <u>même intervalle de variation</u> (en nombre quelconque d'ailleurs).

NOTE BIBLIOGRAPHIQUE

1) - LA COURBE DE CONCENTRATION -

C.GINI - Intorno alle curve di concentrazione - Metron IX. n° 3/4. 1932.

GALVANI - Sulle curve di concentrazione Metron X. n°3, 1932.

Ce dernier donne à son tour les références suivantes :

M.O. LORENZ - Methods of measuring the concentration of wealth Publication of Amer. Stat. Assoc. n° 70, 1905).

E. CHATELAIN - Les successions déclarées en 1905 (Revue Politique et parlementaire, 1907).

J. SEAILLES - La répartition des fortunes en France (Alcan. 1910).

C. GINI - Sulla misura della concentrazione a della variabilità dei caratteri (Atti del R. Inst. Ven. di S.I.A) 1913-1914.

M. HUBER - Eléments de technique statistique - 1943 - pp. 66-68.

M.G. KENDALL - The advanced theory of statistics - tome I. pp. 43-44. 3e édition 1947).

Voir aussi :

P. THIONET - L'école Moderne des statisticiens italiens (2p) Journal de la Société de Statistique de Paris - Janvier-Mars 1946 - p. 23-24.

2) - LES PROBLEMES D'AGREGATION EN ECONOMETRIE

Citons les plus récents travaux seulement après L.KLEIN, THEIL, etc

E. MALINVAUD - L'agrégation dans les modèles économétriques (cahiers du séminaire d'Econométrie 1955).

Aggregations problems in input output models (Proceedings of an international conference on input output analysis, Varenna, 1954).

A. NATAF - Sur des questions d'agrégation en économétrie (Thèse pour le Doctorat es-Sciences - Pub. de l'Institut de Statistique - Volume II; Fasc. 4. 1953).

Voir aussi :

PRAIS et HOUTHAKKER - The analysis of family budget - Cambridge 1955, page 13.

3) - LES LOIS DE DISTRIBUTIONS ECONOMIQUES

G. DARMOIS - Statistique et applications (1934) - p. 114.

M. FRECHET - Sur les formules de répartition des revenus. Revue de l'Inst. Int. Stat. (7.1) 1939.

R. GIBRAT - Une loi des répartitions économiques. Bull. Stat. générale de la France. 1930.

- Les inégalités économiques - Paris édit. Sirey - 1931.

M. HUBER - Eléments de technique statistique - 1943 - (p.122-137-139).

M. CRAMER - Mathematical Methods of Statistics, Princeton 1946 - p. 220-248.

4) - REFERENCES CONCERNANT L'AGREGATION DES LOIS DE DISTRIBUTION

Le théorème de M. FRECHET dont il est ici question était encore inédit ; M. Fréchet nous avait confié ses notes manuscrites. Il emploie l'expression "mélange de populations" (qui est synonyme d'agrégation). De même M. FONSAGRIVE n'a encore rien publié sur ses recherches.

M. HUBER - Eléments de Technique Statistique, 1943 - parle du problème du mélange des populations (page 134). Il cite à ce propos des travaux très anciens :

Alphonse BERTILLON et sa courbe à 2 sommets pour la distribution des tailles des conscrits du Doubs en 1951-1859, décomposée en 2 courbes.

Karl PEARSON et la courbe des décès par âge décomposée en 5 courbes normales.

La question du mélange des populations est mentionnée dans tous les traités de statistique antérieurs à 1940.

Par exemple :

Udny YULE - An introduction to the theory of statistics (11e ed. 1937). p. 103, Complex distributions.

RISSER et TRAYNARD - Les principes de la statistique mathématique (traité de calcul des probabilités de M.E. BOREL, 1933, le partie, chapitre V - Partage des courbes de fréquence.

En revanche KENDALL (1947) la passe sous silence. Et CRAMER (Mathematical methods of statistics, 1946) mentionne seulement (page 56) le cas où on mélange deux distributions, l'une continue, l'autre discontinue. Ce problème de statistique n'est donc redevenu actuel que comme cas le plus simple d'un problème général d'économétrie.[1]

Nota Bene : Il y a lieu de ne pas confondre le mélange des populations avec le produit de plusieurs lois de probabilité, c'est-à-dire l'addition de plusieurs variables aléatoires, question passée au contraire au premier plan du calcul des probabilités.

(1) Exemple d'article récent où il est question de mélange de population : H. MILLER. Elements of symmetry in the skewed income curve (J. Am. Stat. Assoc. March 1955-p. 55)

LES VALEURS MÉDIANES ET LA THÉORIE DE LA MESURE

par

Carlo BONFERRONI

SOMMAIRE :

Moyennes et médianes pour des valeurs discontinues; médiane d'une série et convergence "en médiane". Fonction déterminée dans un intervalle fini; ses "permutations" définies moyennant les "mesures" de certains ensembles ; permutations "au sens étroit". Définition de la "fonctionnelle symétrique" ; l'intégrale et la moyenne arithmétique sont des fonctionnelles symétriques. La permutation croissante et la permutation décroissante ; détermination de la médiane, des terciles, quartiles, etc ... Propriété de minimum de la médiane. Moyenne pondérée de la fonction; pondération à attribuer à la permutation croissante - Une moyenne pondérée est une fonctionnelle symétrique. Permutation croissante et valeurs médianes pour une fonction de deux variables ou plus.

1 - Beaucoup de valeurs moyennes de n observations $x_1, .., x_n$ sont exprimées par une fonction $f(x_1, ..., x_n)$ qui est une fonction analytique et qui pour cette raison se prête facilement à des transformations mathématiques. Ceci n'a pas lieu pour les valeurs "médianes" (médiane, terciles, quartiles, etc ...) bien qu'elles aient de remarquables applications en statistique et ceci dépend évidemment du fait qu'elles sont déterminées après une opération préliminaire de rangement (par ordre croissant ou décroissant) qui ne correspond pas à une fonction analytique, et même, ne se prête pas à une représentation symbolique systématique dans notre construction mathématique actuelle. Il s'ensuit que les problèmes relatifs aux valeurs médianes présentent des difficultés particulières et exigent des procédés qui s'écartent souvent notablement des procédés courants.

Je m'occuperai, en particulier, de la détermination des valeurs médianes quand le nombre des éléments $x_1, ..., x_n$ augmente indéfiniment. Ce problème déjà résolu pour de nombreux types de moyennes, mérite aussi d'être étudié pour les valeurs médianes non seulement à cause des applications pratiques, mais aussi en raison des tendances naturelles de l'esprit mathématique.

2 - Les moyennes les plus importantes acquièrent une signification statistique parce qu'elles résolvent le problème de l'égale répartition (équirépartition d'une certaine grandeur) entre n sujets. La formule mathé-

---

(1) Traduction de la Conférence "I valori e la teoria della misura" parue dans la revue "Giornale di Matematica finanziara" anno XXXVII - Vol I - 1955 - pp 89-153.

matique générale de ce concept (1) conduit à établir que si la grandeur a respectivement les valeurs $x_1, \ldots, x_n$ pour chacun des sujets considérés et la valeur $f(x_1, \ldots, x_n)$ pour l'ensemble qu'ils constituent, la moyenne et la valeur commune μ_n qui conduit à la même valeur collective, c'est-à-dire qui satisfait à la condition :

(1) $$f(\mu_n, \ldots, \mu_n) = f(x_1, \ldots, x_n)$$

dont on pourra tirer (sous certaines conditions concernant la fonction f)

(2) $$\mu_n = g(x_1, \ldots, x_n).$$

Si f est symétrique par rapport aux x_s, g le sera aussi et l'on pourra dire que μ_n est la moyenne symétrique des x_s. Dans les cas plus courants et importants f a la même forme.

(3) $$f(x_1, \ldots, x_n) = \varphi(x_1) + \ldots + \varphi(x_n) :$$

les x_s sont soumis à la même opération φ et les résultats sont additionnés. Les moyennes qui en dérivent peuvent être dites de type arithmétique; elles sont évidemment symétriques.

La médiane ne résout pas un problème d'équirépartition; elle aussi pourtant est une moyenne symétrique, parce que, quel que soit l'ordre suivant lequel les x_s sont donnés, elle est déterminée à la suite d'un rangement par valeurs croissantes (ou décroissantes).

3 - Considérons d'abord une suite infinie $x_1, x_2, \ldots$ par la moyenne de tous les x_s on est amené à définir la limite (si elle existe).

$$\mu = \lim_{n, \infty} \mu_n$$

μ_n étant calculé au moyen des n premiers éléments $x_1, \ldots, x_n$. Si en changeant l'ordre des x_s ne varie pas, celle-ci est une moyenne symétrique de l'infinité des valeurs x_s dépendant de leur ensemble et non de leur ordre de succession. Si μ_n est calculée au moyen de la formule (3) on aura

$$\lim \varphi(\mu_n) = \lim \frac{\varphi(x_1) + \ldots + \varphi(x_n)}{n} = \lim \varphi(x_n) \qquad \text{d'où}$$

Si $\lim x_n$ existe et si φ est une fonction continue.

$$\lim \mu_n = \lim x_n.$$

Dans ces hypothèses la moyenne est donc symétrique car un ordre différent des x_s ne change pas $\lim x_n$. Il se peut toutefois que $\lim \mu_n$ existe sans qu'existe $\lim x_n$; on a alors la limite de x_n en moyenne (de type arithmétique sur la base de φ) ; en particulier en moyenne arithmétique, quadratique, harmonique, géométrique, etc ...

4 - On peut donner une définition analogue pour la médiane en prenant la $\lim \mu_n$, avec μ_n médiane des premières valeurs $x_1, \ldots, x_n$. Si $\lim x_n = x$ existe on aura $\lim \mu_n = x$: en effet étant donné $\varepsilon > 0$ pour n plus grand qu'une certaine valeur x_0, on aura $|x_n - x| < \varepsilon$ et par suite, pour $n > 2 n_0$ on aura $|\mu_n - x| < \varepsilon$. Dans ce cas la médiane est une fonction symétrique de l'infinité des x_s. Mais il peut arriver aussi que la médiane μ_n tende vers une limite alors que x_n ne tend pas vers une limite : $\lim \mu_n$ est alors la limite en médiane de x_n. Par exemple la série indéterminée

$$0, 0, 2, 0, 2, 0, 2, 0, 2, \ldots$$

---

(1) J'ai exposé cette idée en 1923 cf ma note : A proposito di espressioni generali per le medie, Giornale degli Economisti - Mai 1937.

a la limite 1 en moyenne et la limite 0 en médiane. En supprimant le premier zéro, la nouvelle série garde la limite 1 en moyenne mais n'a pas de limite, en médiane. Par contre, la série indéterminée

$$0, 0, 2, 0, -2, 0, 6, 0, -6, 0, 10, 0, -10, 0, 14, 0, -14, \ldots$$

n'a pas de limite en moyenne mais la limite 0 en médiane; si on supprime le zéro initial elle n'a pas de limite ni en moyenne ni en médiane.

5 - Le nombre des x_s devient infini en posant aussi sa fonction x fonction de la variable t dans l'intervalle a - b : la fonction f, qui définit la moyenne devient fonction de la ligne $y = x(t)$, c'est-à-dire une fonctionnelle $F\,x(t)$. La moyenne correspondante s'obtiendra en posant

(4) $$F\mu \quad F\,x(t) \qquad (\mu = \text{constante})$$

Par exemple, si F est l'intégrale de x(t) dans l'intervalle a - b, on aura

$$\int_a^b \mu dt = \int_a^b x(t)\,dt\ , \qquad \mu = \frac{1}{b-a}\int_a^b x(t)\,dt$$

c'est-à-dire la <u>moyenne arithmétique</u> des valeurs prises par x(t). Si F est l'intégrale de x(t) p (t) on obtiendra la <u>moyenne arithmétique pondérée</u> de x(t) avec un poids p(t)dt (élémentaire). En général on pourra de la formule (4) déduire comme fonctionnelle de x(t). Nous nous proposons avant tout d'étendre à ce cas la définition de moyenne <u>symétrique</u>.

6 - Permutation d'une fonction -

Puisque la moyenne est une fonctionnelle de x(t) nous devons déterminer quand une fonctionnelle est symétrique par rapport à la fonction x(t), c'est-à-dire ne varie pas si on substitue à x(t) une fonction $x_1(t)$ représentant un simple changement dans l'ordre de l'infinité des valeurs prises par x(t) dans l'intervalle a - b, c'est-à-dire en somme qui soit une permutation de x(t). Pour une série x_1, x_2, ..., modifier l'ordre des termes signifie considérer une nouvelle série contenant les mêmes termes répétés le même <u>nombre de fois</u> (fini ou infini) cependant, cette condition simple ne peut pas s'appliquer à une fonction x(t) parce qu'une certaine valeur y peut être prise par x(t) en tous les points, par exemple, d'un <u>plurintervalle</u> (intervalle composé d'un nombre fini ou infini d'intervalles disjoints). Il est toutefois naturel dans ce cas de considérer la <u>longueur</u> du plurintervalle (obtenue par l'addition, finie ou en série, des longueurs des intervalles partiels). Seulement, le recours aux plurintervalles n'est pas toujours suffisant, car il existe des ensembles qui ne sont pas des plurintervalles.

Un concept plus général est donc nécessaire : il nous est offert par la <u>mesure</u> d'un ensemble, selon la fameuse définition de Lebesgue.

En indiquant par mes $I(x \leqslant y)$ la mesure (supposée existante) de l'ensemble occupé par les points t qui font $x(t) \leqslant y$, et de façon analogue $x_1(t)$, il est nécessaire d'établir que $x_1(t)$ est une permutation de x(t) si l'on a quel que soit y

(5) $$\text{mes } I\,(x \leqslant y) = \text{mes } I\,(x_1 \leqslant y)$$

Il en découle aussitôt

(6) $$\text{mes } I\,(x = y) = \text{mes } I\,(x_1 = y)$$

En effet ceci peut s'écrire

$$\text{mes } I\,(x \leqslant y) - \lim \text{mes } I\,(x \leqslant z) =$$
$$\text{mes } I\,(x_1 \leqslant y) - \lim \text{mes } I\,(x_1 \leqslant z)$$

où les limites sont calculées pour z croissant et tendant vers y et où elles existent parce que mes $I(x \leq y)$ et mes $I(x_1 \leq z)$ sont fonctions des monotones (non décroissantes) de z.

Il est utile d'observer que l'égalité (6) ne suffit pas à définir une permutation (comme on pourrait au premier moment le supposer) parce qu'elle n'entraine pas l'égalité (5) pourtant nécessaire : on peut en donner un exemple avec les deux fonctions $x(t)$, $x_1(t)$ croissantes dans l'intervalle a-b de A à B de façon que par exemple $x(t) > x_1(t)$ - c'est pourquoi nous prendrons l'égalité (5) pour définir les variations de $x(t)$. L'égalité (5) n'oublie pourtant pas $x_1(t)$ à prendre les mêmes valeurs que $x(t)$. Si en effet on transforme arbitrairement $x(t)$ en un ensemble de mesure nulle, la nouvelle fonction $x_1(t)$ vérifie l'égalité (5). Si on ajoute alors la condition que $x_1(t)$ et $x(t)$ assument les mêmes valeurs et plus précisément que $I(x = y)$ et $I(x_1 = y)$ aient la même mesure (ou "puissance", c'est-à-dire peuvent être mis en relation biunivoque), on obtiendra une permutation au sens étroit. Pourtant par la suite, les propriétés établies seront valables pour n'importe quelle permutation.

7 - Fonctionnelle symétrique -

De même qu'une fonction de $x_1, \ldots, x_n$ est symétrique si elle ne varie pas quand on change l'ordre des variables, de même, par extension naturelle, je dirais qu'une fonctionnelle de $x(t)$ est symétrique si elle reste invariée quand on substitue à $x(t)$ une de ses permutations. Et la fonctionnelle sera seulement symétrique au sens étroit si elle ne varie pas pour une permutation au sens étroit (mais peut varier pour d'autres permutations). Par exemple les extrêmes de $x(t)$ en a-b sont fonctionnelles symétriques au sens étroit parce qu'une permutation générique peut prendre des valeurs non comprises entre des extrêmes de $x(t)$.

La symétrie ainsi définie se ramène à une fonctionnelle d'une seule fonction - Pour une fonctionnelle F de n fonctions on pourra revenir au concept ordinaire de symétrie c'est-à-dire à l'invariabilité de F pour un changement d'ordre dans les fonctions; mais si le nombre des fonctions devient infini, on pourra considérer la symétrie par rapport à ces fonctions infinies; et ceci ramènera à des problèmes du type précédemment traité.

8 - La somme de n nombres est une fonction symétrique de ceux-ci ; à cette propriété correspond la suivante : l'intégrale de $x(t)$ est une fonctionnelle symétrique. Il faut naturellement considérer les intégrales dans le sens de Lebesgue, parce qu'une permutation de $x(t)$, et $x(t)$ même pourraient ne pas être intégrables au sens classique.

La démonstration se déduit immédiatement du processus même de la construction de l'intégrale de Lebesgue. En effet si on introduit les mesures

$$m(y) = \text{mes } I(x \leq y) \quad , \quad M(y) = \text{mes } I(x > y)$$

$$\overline{m}(y) = \text{mes } I(x = y) \qquad m(y) + M(y) = b - a$$

et si on indique respectivement par A et B les limites extrêmes inférieure et supérieure de $x(t)$ dans l'intervalle a - b, l'intégrale de Lebesgue est définie, en substance, par une intégrale de Stieljes : plus précisément

$$(7) \qquad \int_a^b x(t)\, dt = \int_{A-\varepsilon}^{B} y \, dm(y) \qquad [\varepsilon > 0 \text{ arbitraires}]$$

Cette intégrale existe, parce que $m(y)$ est monotone. Il résulte de là qu'une permutation au sens étroit ne modifie pas l'intégrale, parce qu'elle ne modifie ni $m(y)$, ni A ni B. Pour démontrer, ensuite, que la propriété demeure valable pour n'importe quelle permutation il faut remplacer A, B

par deux autres quantités. Notons à cet effet que si les valeurs $x(t)>B_o$; se trouvent sur un ensemble de mesure nulle, pour $y>B_o$ on a $M(y) = 0$, $m(y) = b - a$ et par conséquent $dm(y) = 0$; de même si on a $x(t) < A_o$ pour les points t d'un ensemble de moyenne nulle, pour $y < A_o$, on a $m(y) = 0$ et $dm(y) = 0$: il suffit pour cela d'intégrer de $A_o - \varepsilon$ jusqu'à B_o. Considèrons alors la plus petite valeur B_1 de B_o et la plus grande valeur A_1 de A_o, et appelons les respectivement extrêmale supérieure et extrêmale inférieure de $x(t)$ dans l'intervalle a-b : on aura évidemment

$$A \leqslant A_1 \leqslant B_1 \leqslant B$$

et on pourra retenir dans le calcul de l'intégrale par la formule (7) le fait que A et B représentent les extrêmales de $x(t)$ au lieu des extrêmes (1). Il est clair que $x(t)$ et une de ses commutations $x_1(t)$ ont les mêmes extrêmales (sinon les mêmes extrêmes); donc, en vertu de la formule (7) elles ont la même intégrale.

9 - Il est bon d'établir, pour ce qui suit, que l'intégrale de $x(t)$ est aussi une fonctionnelle symétrique. Il suffit de remarquer qu'évidemment les extrêmales de $|x(t)|$ et de $|x_1(t)|$ coincident ;

$$\text{mes } I(\ |x| \leqslant y) = \text{mes } I(-y \leqslant x \leqslant y) = m(y) -$$
$$- m(-g) + \overline{m}\ (-y)$$

et que le dernier membre n'est pas modifié en passant de $x(t)$ à $x_1(t)$.

10 - Nous devons ajouter et nous en tirerons parti plus loin que la fonction $m(y)$ est continue à droite et discontinue à gauche seulement si $\overline{m}(y) > 0$, avec un écart égal à $\overline{m}(y)$. En effet si on désigne par $m(y+)$, et $m(y-)$ les limites à droite et à gauche de y, il résulte de fait que $m(y)$ tient compte aussi des points t qui rendent $x(t) = y$:

$$m(y+) - m(y) = 0\ , \quad m(y) - m(y-) = \overline{m}(y)$$

Puisque $M(y) = b - a - m(y)$, $M(y)$ aussi se comporte comme $m(y)$.

11 - La permutation croissante -

Si $x(t)$ est monotone (croissante ou décroissante) la médiane est évidemment la valeur de $x(t)$ au point moyen de l'intervalle a - b; les terciles, quartiles, etc ... s'obtiennent de la même manière, endivisant l'intervalle en 3, 4, ..., parties égales. Quand $x(t)$ n'est pas monotone il faudra la transformer en une fonction monotone c'est-à-dire construire la <u>permutation croissante</u> $x_c(t)$, ou <u>bien décroissante</u> $x_d(t)$ = c'est sur elles ensuite qu'on déterminera les valeurs médianes.

Pour avoir $x_c(t)$ nous donnerons la valeur y à tous les points de l'intervalle ayant pour extrêmes

(8) $$t_y = a + m(y)\ , \quad t_y - \overline{m}(y)\ ,$$

intervalle contenu dans celui a - b et de longueur $\overline{m}(y)$. Si $\overline{m}(y) = 0$, les ordonnées du diagramme de $x(t)$ égales à y se rencontrent toutes sur le point t_y, donc, si elles sont deux ou plus, le $x_c(t)$ ne donne pas une permutation au sens étroit le point t_y sera comme $m(y)$ fonction monotone de y, continue à droite; et discontinue à gauche avec l'écart $\overline{m}(y)$ seulement si $\overline{m}(y) = 0$ (voir n° 10) - Le diagramme de $x_c(t)$ vient coincider en correspondance avec l'intervalle dit plus haut, avec un segment parallèle à l'axe t de longueur $\overline{m}(y)$.

(1) Cette simple opération peut souvent être utile (par exemple dans l'application du théorème de la moyenne).

Si on écrit à la place de t_y et si on résout par rapport à y l'équation

$$t = a + m(y) \tag{9}$$

on aura l'équation $y = x_c(t)$ de la permutation croissante, en tenant compte du fait que, lorsque $\overline{m}(y) > 0$, $x_c(t)$ reste constante et = y sur tout l'intervalle (8). Ceci entraine $x_c(t)$, fonction monotone croissante à être aussi une fonction continue. Il faut encore observer que si mes $I(y < x < x + h)$ m(y) reste constante quand y parcourt l'intervalle y - y+h ; la courbe de $x_o(t)$ acquiert alors un règlement parallèle à l'axe y et par suite sur le point t correspondant, la fonction $x_c(t)$ devient plurivalente. Ce fait, comme nous le verrons, a de l'importance pour la détermination de la médiane.

On définit de façon analogue la permutation décroissante x d(t). Son diagramme évidemment, est le symétrique de celui de la $x_c(t)$, par rapport à l'axe du segment a - b.

Un problème d'hydrostatique - et précisément l'équilibre d'un liquide à couches horizontales de diverses densités, qui s'obtient quand les densités se disposent par ordre croissant vers le bas a amené C. SOMIGLIANA (Sulle funzioni reali di una variabile (1) ; sulle funzioni ordinate*) à construire (au moyen de considérations sur les limites) la permutation croissante d'une fonction continue f(x), appelée par lui "fonction ordonnée" de f(x). Dans une lettre à C. SOMIGLIANA, V. VOLTERRA (voir la seconde note citée) étendit le procédé à des fonctions discontinues. L'étude fut reprise par F. SIBIRANI (sulle funzioni ordinatrici (2) qui utilisa le concept de mesure toujours pour des fonctions continues, et indiqua les difficultés que présentent les fonctions discontinues. Dans la présente étude on considère toutes les permutations et en particulier la permutation croissante, sans exclure la possibilité qu'elle soit plurivalente. La statistique examine souvent des fonctions plurivalentes : par exemple une table à double entrée est l'expression au sens mathématique formel, d'une fonction plurivalente, et de son inverse.

12 - Puisque la mesure jusqu'à y, c'est-à-dire m(y), et la mesure pour y, c'est-à-dire $\overline{m}(y)$, sont les mêmes pour x(t) et pour une permutation $x_1(t)$ il est clair que x(t) et $x_1(t)$ conduisent à la même permutation croissante. Soit alors une fonctionnelle F x(t) et F x(t) = F $x_c(t)$ quel que soit x(t) : on aura aussi F $x_1(t)$ = F $x_c(t)$, d'où F x(t) = F $x_1(t)$ et la fonctionnelle sera symétrique. Donc : pour qu'une fonctionnelle soit symétrique, il est nécessaire et suffisant qu'elle ne varie pas lorsqu'on substitue à x(t) sa permutation croissante.

13 - Médiane

Puisque $x_c(t)$ représente x(t) rangé suivant les valeurs croissantes, la médiane est la valeur de $x_c(t)$ située au milieu de l'intervalle a - b. En se reportant au diagramme, on peut dire que les ordonnées qui précèdent l'ordonnée médiane et celles qui la suivent portent sur des ensembles de mesure égale (la moitié de l'intervalle a - b).

Si l'axe du segment a - b rencontre le diagramme en un seul point, il y a une seule valeur médiane ; s'il contient un segment parallèle à l'axe y et faisant partie du diagramme (V n° 11) il y aura tout un intervalle de

(1) SOMIGLIANA - Sulle funzioni reali di una variabile, R. Accademia dei Lincei, 1899, 1° sem.
Sulle funzioni ordinate, id - 2° sem.

(2) SIBIRANI - Sulle funzioni ordinatrici, R. Accademia dei Lincei, 1911, 2° sem; Acc. d. Sc. di Bologna, Mem, 1948.

valeurs médianes. Cependant $x(t)$ ne prend pas toutes ces valeurs, car les ordonnées $x(t)$ comprises dans l'intervalle médian portent le plus souvent sur un ensemble de mesure nulle (souvenons-nous que $m(y)$ reste constante quand y varie dans l'intervalle médian).

Un nombre fini d'éléments peut aussi conduire à un intervalle médian, comme il est bien connu : ce sont pourtant seulement les deux limites extrêmes, de cet intervalle qui appartiennent au groupe d'éléments donné. Si on divise l'intervalle $a - b$ en trois, quatre, ..., parties égales, les ordonnées aux points de division donneront les deux terciles, les trois quartiles, ... ; et un intervalle tercile, un intervalle-quartile, ..., pourront être formés. Si on construit aussi le diagramme de $x_d(t)$; la médiane (unique ou intervalle) constitue l'intersection des deux courbes de $x_c(t)$ et de $x_d(t)$.

14 - Propriété de minimum. Dans le cas discontinu, il est pour ainsi dire évident que la médiane rend minimale la somme des écarts absolus. S'il y a un intervalle médian, ceci a lieu pour tous ses points (1). Comme on peut le prévoir, cette propriété se vérifie aussi dans le cas continu.

Soit en effet,

$$S(z) = \int_a^b |x(t) - z| \, dt$$

la somme (intégrale) des écarts absolus de $x(t)$ pour z; puisqu'une permutation de $x(t)$ ne change pas l'intégrale de $x(t) - z$ ni même celle de $|x(t) - z|$ (v. n° 9), nous pouvons substituer $x_c(t)$ à $x(t)$.

Comme $x_c(t)$ est monotone, croissante, nous aurons :

$$S(z) = \int_A^z m(y) \, dy + \int_z^B M(y) \, dy$$

Les dérivées droite et gauche de $S(z)$ sont (2), étant donné $H(z) = m(z) - M(z)$ (fonction monotone croissante).

$$S'(z+) = H(z) \quad , \quad S'(z-) = H(z-)$$

Si z précède la médiane (ou les médianes) on a $H(z) < 0$; et aussi $H(z-) < 0$ parce que $H(z-) \leqslant H(z)$. Si z est supérieur à la médiane, il résulte au contraire $H(z) > 0$; et aussi $H(z-) > 0$ parce que, pour y compris entre z et la médiane, on a $H(y) > 0$ et croissant. Les dérivées droite et gauche étant négatives avant la médiane, positives après et nulles dans l'intervalle médian (s'il existe, dans cet intervalle : $m = M$) on conclut que $S(z)$ diminue quand z croit jusqu'à la médiane, que Sz reste constant quand z parcourt l'intervalle médian (s'il y en a un) et ensuite croit avec z.

15 - Fonction pondérée :

Supposons que l'on attribue à $x(t)$ un poids (élémentaire) $p(t)dt$ et formons la moyenne pondérée

$$M_p = \int_a^b x(t) \, p(t) \, dt \Big/ \int_a^b p(t) \, dt \; :$$

(1) - $X_1 \ldots x_n$ étant donnés par ordre croissant, la somme des deux écarts absolus de z pour x_1 et x_n est minimale quand $x_1 < z < x_n$; de même, quand $x_2 < z < x_{n-1}$ la somme des écarts entre x_2 et x_{n-1}, etc ... est minimale. Ceci conduit à placer z sur la médiane ou bien sur l'intervalle médian.

(2) On démontre facilement que si $F(z)$ a une limite droite (gauche) en z, cette limite est la dérivée droite (gauche) en z de l'intégrale de F en $a^- z$. On se souviendra aussi que $m(y)$ et $M(y)$ sont continues à droite (n° 10).

nous nous proposons de vérifier qu'une permutation de $x(t)$ laisse M_P inchangée. Cette propriété, évidente pour le cas discontinu - parce que toutes les valeurs x_s portent avec elles, dans un déplacement, leur propre pondération p_s, on peut faire ensuite la somme des coefficients de pondération de valeurs successives égales - n'est pas évidente dans le cas continu, car il est nécessaire avant tout d'établir quelle est la pondération $p_c(t)dt$ à donner à la permutation $x_1(t)$.

Nous commencerons par étudier la permutation croissante $x_c(t)$ en déterminant le poids $p_c(t)dt$ qui lui est attaché. Sur le segment $a^- t$, $x_c(s)$ absorbe l'ensemble $I(x<y)$, car $y = x_c(t)$: c'est-à-dire dans l'hypothèse $\overline{m}(y) = 0$. Mais si $\overline{m}(y) > 0$, il en résulte que $x_c(t)$ est constante $= y$ dans un intervalle $\alpha - \beta$ de longueur $\overline{m}(y)$; le segment $a^-\alpha$ contient l'ensemble $I(x<y)$ et le segment $a^- \beta$, ce même ensemble et en plus $I(x = y)$. C'est pourquoi, pour $\alpha \leq t \leq \beta$ nous attribuerons au segment $a^- t$ le poids total de $I(x < y)$ plus une partie du poids de $I(x=y)$, proportionnelle à $t - \alpha$ (1)= c'est-à-dire nous poserons :

$$(10) \qquad \int_a^t P_c(s)ds = \int_{x<y} p(s)\, ds + \frac{t - \alpha}{\beta - \alpha} \cdot \int_{x=y} p(s)\, ds \quad \left[y = x_c(t) \right]$$

Pour $t = b$, on a évidemment (B représentant l'extrême supérieure de $x(t)$ dans l'intervalle $a^- b$)

$$(11) \qquad \int_a^b p_c(s)ds = \int_{x<B} p(s)ds + \int_{x=B} p(s)ds = \int_a^b p(s)\, ds$$

c'est-à-dire que le poids total n'a pas varié. Si $\overline{m}(y) = 0$, le dernier terme du deuxième membre de l'équation (10) disparait.

16 - Pour déterminer $p_c(t)$ il est essentiel de noter que le deuxième membre de l'équation (10) est une fonction $\varphi(t)$ <u>absolument continue</u> du moins dans l'hypothèse - où nous nous plaçons - que $p(t)$ est fonction <u>limitée</u> en $a^- b$ c'est-à-dire que l'on a $p(t) < P$, P étant indépendant de t. Je montrerai précisément que le rapport d'accroissement de $\varphi(t)$ dans n'importe quel intervalle de $a^- b$ est inférieur à P ; et ceci - comme il est bien connu - suffira pour qu'on affirme que $\varphi(t)$ est absolument continue. L'intervalle $t^- t+h$ $(h > 0)$ étant fixé, supposons pour nous placer dans des conditions générales que t porte sur un segment $\alpha_1 - \beta_1$, dans lequel, $x_c(t)$ = constante $= y_1$ et $t+h$ dans $\alpha_2 - \beta_2$ où $x_c(t)$ = constante $= y_2$; et posons :

$$h_1 = \beta_1 - t, \; h_2 = \alpha_2 - \beta_2, \; h_3 = t+h - \alpha_2.$$

Si t ne se trouve pas sur un segment $\alpha_1 - \beta_1$, $h_1 = 0$; et si $t+h$ n'appartient pas à $\alpha_2 - \beta_2$, $h_3 = 0$, puisque quant t varie dans $\alpha_i \beta_i$, seul le deuxième terme de $\varphi(t)$ varie, nous aurons :

$$(12) \quad \begin{cases} \varphi(\beta_1) - \varphi(t) = \dfrac{h_1}{\beta_1 - \alpha_1} \displaystyle\int_{x=y_1} p(s)\, ds = h_1 p_1 \\ \varphi(\alpha_2) - \varphi(\beta_1) = \displaystyle\int_{x<y_2} p(s)\, ds - \int_{x \leq y_1} p(s)ds = \int_{y_1<x<y_2} p(s)ds = h_2 p_2 \\ \varphi(t+h) - \varphi(\alpha_2) = \dfrac{h_3}{\beta_2 - \alpha_2} \displaystyle\int_{x=y_2} p(s)ds = h_3 p_3 \end{cases}$$

Posons :

p_1 = valeur moyenne de $p(s)$ en $I(x=y_1)$

(1) On pourrait bien entendu faire une autre convention.

(12) p_2 = valeur moyenne de p(s) en $I(y_1 < x < y_2)$

p_3 = valeur moyenne de p(s) en $I(x=y_2)$

Puisque $\varphi(t)$ est évidemment fonction croissante de t, les accroissements (12) expriment les variations. En ajoutant les expressions (12), on obtient :

(13) $$\varphi(t+h) - \varphi(t) = h_1p_1 + h_2p_2 + h_3p_3 = h\,\overline{p} < h\,P$$

$\overline{p}$ étant la moyenne de p_1, p_2, p_3, et par conséquent, moyenne de p(t) en $I(y_1 \leqslant x \leqslant y_2)$. Ceci démontre la continuité absolue de φ (t).

Par suite, la dérivée $\varphi'(t)$ existe dans l'intervalle a⁻b, sauf à la rigueur pour un ensemble de mesure nulle; et l'intégrale de $\varphi'(t)$ est égale à $\varphi(b) - \varphi(a)$, c'est-à-dire dans notre cas, $\varphi(b)$, puisque $\varphi(a) = 0$. Nous poserons alors $p_c(t) = \varphi'(t)$ dans les points de dérivabilité. Quelle que soit la valeur que l'on donne à $p_c(t)$ sur les points qui restent, on aura toujours

$$\int_a^t p_c(s)ds = \int_a^t \varphi'(s)ds = \varphi(t)$$

et l'équation (10) sera satisfaite. Ayant ainsi obtenu une solution pour $p_c(t)$ on pourra en déduire d'autres en alternant $p_c(t)$ dans un ensemble de mesure nulle.

17 - Quant t est intérieur à un intervalle $\alpha - \beta$ où $x_c(t)$ est constante et égale y, la dérivée $\varphi'(t)$ se détermine immédiatement, en notant que $\varphi(t)$ dans $\alpha - \beta$ est une fonction linéaire, représentée par le deuxième terme du second membre de l'équation (10). On a ainsi :

(14) $$p_c(t) = \frac{1}{\beta - \alpha} \int_{x=y} p(s)ds = \text{constante} = \overline{p}$$

avec $\overline{p}$ moyenne arithmétique de p(s) dans l'ensemble I(x = y) qui a pour mesure $\overline{m}(y) = \beta - \alpha$

La coincidence de $\varphi'(t)$ coincide avec une valeur moyenne de p(s) subsiste même si t est en dehors d'un segment $\alpha - \beta$, dans l'hypothèse où p(t) est une fonction continue. En effet, l'équation (13) donne :

$$\frac{\varphi(t+h) - \varphi(t)}{h} = \overline{p},$$

$\overline{p}$ = valeur moyenne de p(s) dans l'intervalle $I_h(y \leqslant x \leqslant y\ )$ avec $y = x_c(t)$, $y_h = x_c(t+h)$. Quand h tend vers 0, y_h tend vers y, et l'ensemble I_h tend vers l'ensemble I(x=y) complété par son dérivé (1). Si p(t) est fonction continue, il est clair que p moyenne dans I_h, tend vers une valeur moyenne dans l'ensemble limite de I_h, formé, comme on l'a dit, de I(x = y) et de son dérivé. Mais la valeur $p(t_o)$ en un point d'accumulation de t, quand t varie dans un ensemble I, est elle aussi - en raison de la continuité de p(t) - une valeur moyenne entre celles que p(t) assume dans I = c'est pourquoi la limite de $\overline{p}$, c'est-à-dire $\varphi'(t)$, est une moyenne de p(t) dans l'ensemble I(x = y).

---

(1) J'applique ici, la définition de limite et de continuité d'un ensemble variable que j'ai exposé récemment (Alcune proprieta generali di un insieme variabile, Boll. Unione matem. Ital. 1954, Mars). Si I(t) est un ensemble (par exemple numérique), fonction de t, le point x_o sera "interlimite" de I(t) pour t tendant vers t_o, si, étant donné $\varepsilon > 0$ tout autour de t_o, on trouve au moins un point t pour lequel I(t) possède des éléments compris entre $x_o - \varepsilon$ et $x_o + \varepsilon$. Si on trouve un intervalle entourant t_o tel qu'en chacun de ses points t, cette limitation soit valable, x_o sera "interlimite stable". Les interlimites forment la classe "limite" en t_o, toujours existante et fermée. Si elle contient seulement des interlimites stables, elle est la "limite" de I(t) et on peut dire que I(t) tend vers elle. I(t) est fonction continue de t en t_o si, elle tend vers l'ensemble formé par $I(t_o)$, et par son dérivé.

On a supposé que $\varphi'(t)$ existait en t. Mais quand elle n'existe pas on peut attribuer à $p_c(t)$, comme on l'a déjà dit(n° 16), une valeur arbitraire : nous donnerons encore à $p_c(t)$ une valeur moyenne entre celles que $p(t)$ assume en $I(x = y)$ (par exemple la demi-somme des extrêmes); que nous pouvons donc, conclure en général que si la densité de pondération $p(t)$ est fonction continue de t, on peut satisfaire à la condition (10) avec une densité $p_c(t)$ qui est une moyenne des densités $p(s)$ sur les points qui rendent $x(s) = x_c(t)$.

Si, en particulier, $p(t)$ = constant = 1, on pourra aussi poser $p_c(t) = 1$.

18 - La structure de la moyenne $p_c(t)$ peut être précisée dans les cas plus courants, correspondant aux hypothèses que nous allons maintenant indiquer.

Supposons que $x(t)$ aussi soit fonction continue : y une fois fixé, si en un point on a $x(s) < y$, il existera tout autour de s un intervalle dans lequel $x(s)$ est encore $< y$, et de ce fait, l'intervalle $I(x < y)$ se composera d'intervalles (en nombre même infini) disjoints, qui pourront être disposés en série $t_1 — t_2$, $t_3 — t_4$, ...

Si $\overline{m}(y) > 0$, un segment α - β est formé dans lequel $x_c(t)$ est constant = y et comme on l'a vu (n° 17) $p_c(t)$ est la moyenne arithmétique simple de $p(s)$ en $I(x = y)$.

Si, en revanche, $\overline{m}(y) = 0$, le point t où $x_c(t) = y$ est (comme on l'a vu au n° 11).

$$t = a + \text{mes}\ I(x < y) = a + \sum_i (t_{2i} - t_{2i-1}) \tag{15}$$

Faisons, maintenant, varier t dans un intervalle qui l'entoure σ, et supposons - comme il advient généralement - que l'on ait toujours $\overline{m}(y) = 0$: les points t_i varieront en conséquence, et en particulier t_{2i} croitra avec t, tandis que t_{2i-1} décroitra. Notons que si y représente un maximum (relatif) de $x(s)$, deux points t_{2i-1}, t_{2i} coïncident et cessent d'exister si t augmente : il faudra, alors, prendre une marge σ seulement à gauche ; et, de la façon analogue, seulement à droite si y est un minimum de $x(t)$. Si on prend σ par le côté opposé on pourra avoir un résultat différent (comme l'exemple au n° 22 le confirmera).

En supposant que $x(t)$ soit une fonction dérivable, on aura $y = x(t_i)$, fonction dérivable de t_i ; donc, dans σ, t_i sera aussi dérivable par rapport à y ; par conséquent, en vertu de l'équation (15), t sera une fonction dérivable de y (en supposant que si la somme devient une série, elle sera dérivable terme par terme) ; et enfin, y sera dérivable par rapport à t. Mais alors, t_i aussi sera une fonction dérivable de t ; et l'expression (15) pourra donc être dérivée par rapport à y comme par rapport à t. En général, nous pouvons écrire (en remarquant que t_{2i} croît et que t_{2i-1} décroît quand t croît)

$$dt = \sum (dt_{2i} - dt_{2i-1}) = \sum |dt_i| \tag{16}$$

Si on divise par dt et si on indique par t'_i la dérivée par rapport à t, on a :

$$1 = \sum |t'_i| \tag{17}$$

Si, au contraire, on divise par dy, on obtient la relation :

$$\frac{1}{x'_c(t)} = \sum \frac{1}{|x'(t_i)|} \tag{18}$$

qui relie la dérivée en t de la permutation aux dérivées de x(s) aux points t_i correspondant à t. D'après ce qui a été dit, si $y = x_c(t)$ est un maximum ou un minimum de x(s), les relations (17) et (18) seront valables à droite de t pour un certain groupe de t_i, et à gauche pour un autre groupe : et si t_j appartient à deux groupes à la fois, la fonction $t_j(t)$ variera (en général) en passant d'un groupe à l'autre. (V. l'exemple au n° 22).

19 - Venons-en à la détermination de $p_c(t)$. L'équation (10) devient, en remarquant que le deuxième terme du deuxième membre manque parce que $\overline{m}(y) = 0$ en σ et que $I(x < y)$ est plurintervalle :

$$(19) \qquad \int_a^t p_c(s)ds = \sum_i \int_{t_{2i-1}}^{t_{2i}} p(s)\,ds$$

On trouve en la différenciant :

$$(20) \qquad p_c(t)\,dt = \sum_i \left[p(t_{2i})\,dt_{2i} - p(t_{2i-1})\,dt_{2i-1}\right] = \sum p(t_i)\,\left|dt_i\right|$$

c'est-à-dire le poids (élémentaire) pour $x_c(t)$ est la somme des poids aux points t_i où $x(t_i) = x_c(t)$.

Ainsi, dans le cas continu, une propriété évidente pour les distributions discontinues (v.n°15) reste valable. En divisant ensuite l'expression (20) par dt on obtient la solution continue

$$(21) \qquad p_c(t) = \sum p(t_i)\,\left|\,t'_i\,\right| \quad ;$$

celle-ci, compte tenu de la relation (17), exprime que la densité (continue) de pondération à attribuer à $x_c(t)$ est moyenne pondérée de la densité aux points t_i correspondant à t, avec comme poids $|t'_i|$ pour $p(t_i)$. La nature de la moyenne $p_c(t)$ se trouve ainsi précisée dans les hypothèses considérées.

20 - Montrons maintenant que l'intégrale de x(s) p(s) coïncide avec celle de $x_c(t)\,p_c(t)$ dans l'intervalle $a^- b$.

Commençons par considérer l'ensemble U des segments α_i——β , dans lesquels $x_c(t)$ = constante = y_i. Dans α_i——β_i =

$$p_c(t) = \frac{1}{\beta_i - \alpha_i} \cdot \int_{x=y_i} p(s)\,ds \quad \left[\overline{m}(y_i) = \beta_i - \alpha_i\right]$$

Comme $x_c(t) = x(s)$ = constante = y_i quel que soit s dans $I_i(x = y_i)$, si on multiplie le premier membre par $x_c(t)$ et le deuxième par x(s) à l'intérieur de l'intégrale, nous aurons :

$$\int_{\alpha_i}^{\beta_i} x_c(t)\,p(t)\,dt = \int_{x=y_i} x(s)\,p(s)\,ds$$

et, en faisant la somme :

$$(22) \qquad \int_U x_c(t)\,p_c(t)\,dt = \int_I x(s)\,p(s)\,ds$$

où I est la "somme" des ensembles disjoints I_i.

Soit, alors, t extérieur à U, c'est-à-dire situé dans son complément V. Puisque $x_c(t) = x(t_i)$ quelle que soit la valeur correspondante de t_i, si on multiplie par $x_c(t)$ le premier membre de l'équation (20) et par $x(t_i)$ le terme du deuxième membre contenant $p(t_i)$, en intégrant ensuite sur V, on aura :

$$\int_V x_c(t)\,p_c(t)\,dt = \sum_i \int_{\gamma_i} x(t_i)\,p(t_i)\,dt_i$$

où γ_i est l'ensemble parcouru par t_i quand t varie dans V. Puisque t_i est fonction continue de t, croissante ou décroissante, γ_i est un intervalle dont les extrémités sont des points τ pour lesquels $y = x(\tau)$ est maximum ou minimum, ou bien $\overline{m}(y) > 0$.

De tels intervalles γ_i (disjoints) complètent évidemment le complément J de I dans $a \overline{\quad} b$: donc :

$$\int_V x_c(t)\, p_c(t)\, dt = \int_J x(t)\, p(t)\, dt \tag{23}$$

En additionnant (22) et (23) on obtient en définitive, comme on le désirait :

$$\int_a^b x_c(t)\, p_c(t)\, dt = \int_a^b x(t)\, p(t)\, dt \tag{24}$$

21 - Des égalités (24) et (11), il résulte donc, que la moyenne pondérée M_p ne varie pas quand on calcule sur la permutation $x_c(t)$ avec la pondération $p_c(t)$ définie précédemment.

Passons, maintenant, à une autre permutation $x_1(t)$ de $x(t)$. Le poids $p_1(t)dt$ à assigner à $x_1(t)$ devra être tel que, selon les critères habituels, on ait pour résultat, quel que soit y

$$\int_{x \leqslant y} p(s)\, ds = \int_{x_1 \leqslant y} p\,(s)\, ds$$

De là, par un raisonnement analogue à celui employé au n° 6 pour $I(x \leqslant y)$, on déduit :

$$\int_{x=y} p(s)ds = \int_{x_1=y} p_1(s)ds, \qquad \int_{x<y} p(s)ds = \int_{x_1<y} p_1(s)\, ds \tag{25}$$

Ceci posé, construisons la permutation croissante $x_{1c}(t)$ de $x_1(t)$, et déterminons le poids $p_c(t)dt$ à lui assigner. Nous savons déjà (n° 12) que $x_{1c}(t)$ est : $x_c(t)$; de plus, en vertu de la relation (25), le deuxième membre de l'équation (10) ne varie pas quand on substitue $p_1(s)$ à $p(s)$, d'où l'on peut admettre que $p_{1c}(t) = p_c(t)$. Donc, s(t) et $x_1(t)$ ont la même permutation croissante et avec le même poids. Comme les moyennes pondérées de x(t) et $x_1(t)$ coïncident avec les moyennes pondérées de leurs permutations croissantes, et que, celles-ci sont égales, on conclut que les deux moyennes de x(t) et $x_1(t)$ seront aussi égales. En résumé, on peut dire qu'une moyenne pondérée est une fonctionnelle symétrique. Ce résultat ne pouvait se déduire directement de l'observation formulée dans le paragraphe 12, parce que la moyenne pondérée de x(t) n'est pas seulement une fonctionnelle de x(t) mais aussi de p(t) et cette fonction est modifiée quand on passe à une permutation $x_1(t)$.

22 - Exemple :

Appliquons à un exemple - tout-à-fait simple pour éviter des complications de nature exclusivement technique - les procédés exposés. Soit x(t) dans l'intervalle $2 \overline{\quad} 2$, défini comme suit :

Intervalles	$-2 \overline{\ } -1$	$-1 \overline{\quad} 0$	$0 \overline{\quad} 1$	$1 \overline{\quad} 2$
x(t)	$-1-t$	$1+t$	$\frac{2-t}{2}$	$\frac{t}{2}$

La courbe de x(t) est la ligne brisée, ayant pour sommets (-2, 1) (-1, 0) (0, 1) (1, 1/2) (2, 1).

Pour $y < 1/2$ on a seulement 2 points t_1, t_2; et pour $y > 1/2$ on introduit aussi t_3, t_4 (coincident pour $y = 1/2$). Précisément :

$$y < 1/2 , \; t < -1$$

$$t_1 = -1 - y , \; t_2 = -1 + y , \; t_2 - t_1 = 2y \quad t_1 + t_2 = -2$$

$$t = -2 + t_2 - t_1 = -2 + 2y , \; y = x_c(t) = \frac{2+t}{2} \quad x_c \; (t) = 1/2$$

$$t_1 = -2 - t/2 \quad t_2 = t/2 , \; t'_1 = t'_2 = 1/2$$

$$y \geqslant 1/2 , \; t \geqslant -1$$

$$t_1 = -1 - y , \; t_2 = -1 + y \quad t_3 = 2 - 2y , \; t_4 = 2y$$

$$t_2 - t_1 = 2y , \; t_1 + t_2 = -2 ; \; t_4 - t_3 = 2 + 4y , \; t_3 + t_4 = 2$$

$$t = -2 + t_2 - t_1 + t_4 - t_3 = -4 + 6y , \; y = x_c(t) = \frac{4+t}{6} \quad x'_c(t) = 1/6$$

$$t_1 = \frac{-10-t}{6} , t_2 = \frac{-2+t}{6} , t_3 = \frac{2-t}{3} , t_4 = \frac{4+t}{3}$$

$$|t'_1| = t'_2 = 1/6 \qquad |t'_3| = t'_4 = 1/4$$

La médiane est $x_c(0) = 4/6$; les quartiles sont $x_c(-1)$, $x_c(0)$, $x_c(1)$ c'est-à-dire 3/6, 4/6, 5/6. On vérifie que $\sum |t'_i| = 1$. A gauche et à droite de $t = -1$ on a, pour t_1 et t_2 deux expressions différentes en fonction de t ; la dérivée de $x_c(t)$ à gauche est 1/2, à droite 1/6 et ceci parce que pour $t = -1$, $x(t)$ passe par un minimum.

La moyenne (simple) peut être calculée au moyen de $x(t)$ ou bien de $x_c(t)$. En effet :

$$\int_{-2}^{+2} x(t)\,dt = \int_{-2}^{-1} (-1-t)\,dt + \int_{-1}^{0} (1+t)\,dt + \int_{0}^{1} \frac{(2-t)\,dt}{2} + \int_{1}^{2} \frac{t\,dt}{2} =$$

$$= 1/2 + 1/2 + 3/4 + 3/4 = 5/2$$

$$\int_{-2}^{2} x_c(t)\,dt = \int_{-2}^{-1} \frac{2+t}{2}\,dt + \int_{-1}^{2} \frac{4+t}{6}\,dt = 1/4 + 9/4 = 5/2$$

Donc, la moyenne est $\frac{5/2}{4} = 5/8$

Passons aux moyennes pondérées, en supposant $p(t) = 2 + t$ et en calculant $p_c(t)$

$$-2 \;\overline{\qquad}\; -1$$

$$p_c(t) = p(t_1)\,|t'_1| + p(t_2)\,t'_2 = \frac{2 + t_1 + 2 + t_2}{2} = \frac{4-2}{2} = 1$$

$$-1 \;\overline{\qquad}\; 2$$

$$p_c(t) = p(t_1)\,|t'_1| + p(t_2)t'_2 + p(t_3)\,|t'_3| + p(t_4)t'_4 =$$

$$= \frac{2 + t_1 + 2 + t_2}{6} + \frac{2 + t_3 + 2 + t_4}{3} = \frac{4-2}{6} + \frac{4+2}{3} = 7/3$$

Pour $t = -1$ la densité de pondération est 1 à gauche et 7/3 à droite. Vérifions l'invariance du poids total

$$\int_{-2}^{2} p(t)\,dt = \int_{-2}^{2} (2+t)\,dt = 8$$

$$\int_{-2}^{2} p_c(t)\,dt = \int_{-2}^{-1} dt + \int_{-1}^{2} \frac{7}{3}\,dt = 1 + 7 = 8$$

Il en résulte en outre :

$$\int_{-2}^{2} x(t)\,p(t)\,dt = \int_{-2}^{-1} (-1-t)\,(2+t)\,dt + \int_{-1}^{0} (1+t)\,(2+t)\,dt + \int_{0}^{1} \frac{4-t^2}{2}\,dt +$$

$$+ \int_{1}^{2} \frac{t}{2}(2+t)\,dt = 1/6 + 5/6 + 11/6 + 16/6 = 11/2$$

$$\int_{-2}^{2} x_c(t)p_c(t)\,dt = \int_{-2}^{-1} \frac{2+t}{2}\,dt + \int_{1}^{2} \frac{4+t}{6}\,\frac{7}{3}\,dt = 1/4 + 21/4 = 11/2$$

L'une ou l'autre intégrale, divisée par le poids total, donne :

$$M_p = \frac{11/2}{8} = 11/16 .$$

23 - Pour avoir la médiane, les terciles, les quartiles, etc ... de $x(t)$ avec pour poids $p(t)$, il suffira de diviser la surface du diagramme de $y = p_c(t)$ en 2, 3, 4, ... parties égales : si les ordonnées à diviser portent sur les points τ_1, τ_2 ..., les valeurs correspondantes de $x_c(t)$ donnent le résultat cherché. Dans l'exemple précédent, si on a $p_c(t)$ = constante = 1, dans l'intervalle $-2 \overline{\quad\quad} -1$ et $p_c(t)$ = constante = $7/3$ dans l'intervalle $-1 \overline{\quad} -2$, la détermination est très facile. Pour la médiane pondérée, c'est par exemple :

$$T_1 = 2/7 \qquad M_{ep} = \frac{4 + 2/7}{6} = 5/7$$

valeur un peu supérieure à la moyenne pondérée.

24 - Fonction de deux variables.

On peut immédiatement étendre les concepts et les définitions précédentes à une fonction mesurable $x(u,v)$ de deux variables, définie sur les points d'une région plane (R) de mesure (de surface) R. Sur la base de la mesure $m(y)$ de $I(x \leqslant y)$ on définit les permutations $x_1(u,v)$ et en particulier celle que j'appellerai encore permutation croissante et qui sera une fonction $x_c(t)$ de la seule variable t, construite sur l'intervalle $0 \overline{\quad} R$ par le procédé employé pour $x(t)$, c'est-à-dire en posant $x(t) = y$ pour les points de l'intervalle qui a pour extrémités :

$$m(y) \quad , \quad m(y) - \overline{m}(y) \qquad \left[\overline{m}(y) = \text{mes } I(x=y)\right]$$

Il est intéressant de remarquer que de cette façon les ordonnées infinies de la courbe de $x(u,v)$ dans un champ à deux dimensions, sont transférées, dans un champ à une seule dimension, et disposées par ordre de grandeur. La gradation de ces ordonnées, toujours possible quand leur nombre est fini même si elles dépendent de deux indices entiers, peut donc s'effectuer aussi pour une distribution continue à deux variables.

Les permutations auront la même intégrale (de Lebesgue) ; en particulier

$$\iint_{(R)} x(u,v)\,du\,dv = \int_{0}^{R} x_c(t)\,dt,$$

égalité qui peut servir de définition de l'intégrale de Lebesgue et qui constitue en substance, une formule de réduction d'une intégrale double à une intégrale simple.

$x_c(t)$ une fois construite, sa valeur pour R/2 sera la médiane de la double distribution de valeurs $x(u,v)$. Au moyen de $x_c(t)$ on pourra aussi calculer la moyenne de $x(u,v)$ en (R). Pour les moyennes pondérées il faudra déterminer le poids $p_c(t)$ dt pour $x_c(t)$, le poids $p(u,v)$ dudv pour $x(u,v)$, étant connu, en posant une condition analogue à l'équation (10).

Enfin, on peut étendre ce qui précède à des fonctions de trois variables ou plus.

ANALYSES D'ARTICLES PARUS DANS DES REVUES ÉTRANGÈRES

Mme LAFON

SANKHYA

Volume 17 Part. 1-2 - Août 1956

CLASSIFICATION AND ANALYSIS OF LINKED BLOCK DESIGNS

by J. ROY and R.G. LAHA

Un "EQUI REPLICATE INCOMPLETE BLOCK DESIGN" est un arrangement de v variétés en b blocks de k groupes chacun, k v, chaque variété apparaissant au plus une fois dans chaque block, et ensemble dans r blocks.

Cet arrangement est appelé "Linked Block design" si $\mu_{ij} = \mu \ \forall \ i \neq j = i, 2, \ldots, b$ et "Balanced Incomplete Block design" si $\lambda_{ij} = \lambda \ \forall \ i \neq j = 1, 2, \ldots, v$.

Une condition nécessaire et suffisante pour qu'un BIB design soit un LB est que $v = b$.

L'auteur donne ensuite la définition d'un PBIB design à m classes associées.

L'ANALYSE INTRA-BLOCK de LB DESIGN est complètement faite.

L'auteur donne un exemple numérique : $v = 17$, $b = 9$, $k = 8$, $r = 4$, $\mu = 3$.

C - CLASSIFICATION DES LB DESIGNS :

en trois groupes principaux :
1) BIB designs symetriques,
2) PBIB designs,
3) Designs irréguliers.

CONDITION NECESSAIRE ET SUFFISANTE POUR QU'UN PBIB DESIGN A DEUX CLASSES ASSOCIEES SOIT UN LB DESIGN.

C'est que la matrice $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ ait une seule racine $\neq 0$ et qu'elle soit d'ordre $b - 1$ pour la matrice d'association du PBIB design.

Avec $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$,

$$a_{11} = r + \lambda_1 p^1_{11} + \lambda_2 p^1_{12} - \lambda_1 n_1$$

$$a_{12} = \lambda_1 p^1_{11} + \lambda_2 p^2_{12} - \lambda_1 n_1$$

$$a_{21} = \lambda_1 p^1_{21} + \lambda_2 p^1_{22} - \lambda_2 n_2$$

$$a_{22} = r + \lambda_1 p^2_{21} + \lambda_2 p^2_{22} - \lambda_2 n_2$$

LISTE ET PLAN DE TOUS LES LB DESIGNS OU $R, K \leqslant 10$

ON THE RECOVERY OF INTER BLOCK INFORMATION IN VARIETAL TRIALS
by C. RADHAKRISHNA RAO

Commentaires sur la méthode de "Intra and inter block analysis" proposée par l'auteur en 1947.

METHODE Q

On résoud les équations

$$Q_i = \frac{r(k-1)}{k} t_i - \frac{\lambda_{i1}}{k} t_1 - \dots - \frac{\lambda_{iv}}{k} t_v \qquad i = 1, \dots, v$$

avec les notations habituelles.

L'auteur fait l'analyse générale :

Exemples : 1/ BIB
2/ PBIB
3/ LB on introduit ici l'écriture matricielle.

METHODE P

Intra block analyse :

aux équations $Q = (rI - \frac{1}{k} N'N)\underset{\sim}{t}$

on ajoute les équations

$$\underset{\sim}{B} = k\underset{\sim}{b} + N\underset{\sim}{t}$$

finalement on résoud les équations

$$\underset{\sim}{P} = (k - \frac{NN'}{r})\underset{\sim}{b}$$

Intra et inter block analyse :

aux équations $\underset{\sim}{Q} = (wrI - \frac{w - w'}{k} N'N)\underset{\sim}{t}$

on ajoute

$$\underset{\sim}{B} - km\underset{\sim}{u} = k\underset{\sim}{b} + N\underset{\sim}{t}$$

A GENERAL CLASS OF QUASIFACTORIAL AND DESIGNS RELATED
by C. RADHAKRISHNA RAO

DEFINITION D'UN" QUASIFACTORIEL DESIGN" :

Soient v variétés, que l'on identifie avec le système suivant :

$$(x_1, x_2, \dots, x_n)$$

$$x_i = 1, 2, \dots, p_i \qquad i = 1, \dots, n$$

et b blocks de k variétés chacun tels que

1) chaque variété est utilisée r fois

2) les deux variétés représentées par (x_i) et (y_i) apparaissent dans $\lambda_{c_1 \dots c_n}$ blocks avec $c_i = 1$ ou 0 suivant que $x_i = y_i$ ou $x_i \neq y_i$.

$$vr = bk$$

$$r(k-1) = p(c_1, \dots, c_n)$$

$$p(c_1, \dots, c_n) = (p_{m_1} - 1)\, p_{m_2} - 1) \dots (p_{m_s} - 1)$$

avec $c_{m_1}, \dots, c_{m_s} = 1$, les autres = 0.

On considère dans cet article les "quasifactorial designs" de dimension 2.

CONSTRUCTION DE "QUASIFACTORIAL DESIGNS" de DIMENSION DEUX

1) Provenant de carrés latin orthogonaux

a) $v = pq \quad b = q(q-1) \quad k = p \quad r(q-1) \quad \lambda_{01} = \lambda_{10} = 0 \;, \lambda_{11} = 1$

exemples numériques.

b) $v = pq \,,\; b = q^2 \,,\; k = p \quad r = q \,,\; \lambda_{11} = 1 = \lambda_{10} \;, \lambda_{01} = 0$

2) provenant de la réunion de 2 BIB designs de paramètres

$$v = p \,,\; b = b_1 \,,\; r = r_1 \,,\; k \,,\; \lambda_1$$

$$v = q \,,\; b = b_2 \,,\; r = r_2 \,,\; k \,,\; \lambda_2$$

on obtient les designs de paramètres

$$v = px \,,\; b = pb_1 + qb_2 \,,\; r = r_1 + r_2 \,,\; k \,,\; \lambda_{11} = 0 \,,\; \lambda_{01} = \lambda_1 \,,\; \lambda_{10} = \lambda_2$$

DESIGNS A TREILLIS CIRCULAIRE

Soient n cercles concentriques et n diamètres de finissant $2\,n^2$ points, si les cercles et les diamètres sont pris comme blocks on obtient un design de paramètres :

$$v = 2n^2 \,,\; b = 2n \,,\; k = 2n \quad r = 2.$$

l'analyse est faite en utilisant la méthode P.

THE ANNALS OF MATHEMATICAL STATISTICS
Vol. 27 - N° 2 - Juin 1956

Melle A. M. GERVAISE

ALTERNATIVES MODELS FOR THE ANALYSIS OF VARIANCE BY M. SCHEFFE -

L'auteur passe en revue les principaux modèles mathématiques qui font l'objet de l'analyse de la variance, leur formulation, justification et l'inférence statistique qu'ils permettent; cet article est semblable à celui de S.L. CRUMP (1951) "The present status of variance composant analysis" - Biometrica - Vol. 7; pp. 1 à 16 - Toutefois il ne le répète pas. Dans ces deux articles les modèles particuliers et difficiles né sont pas traités; pour ce qui concerne ceux relatifs à l'industrie et à la physique, l'auteur renvoie à Kempthorne et Wilk (1953-1955) "A series of reports beginning.December 1953, on models for the analysis of variance, Analysis of variance project, Statistical Laboratory, Iowa state college -

THE THEORY OF DECISION PROCEDURES FOR DISTRIBUTIONS WITH MONOTONE LIKEHOOD RATIO - By S. KARLIN and H. RUBIN

Les auteurs généralisent et précisent les résultats de S.G. ALLEN "Annals of Mathematic Statistic Vol. 24 (1953) pp. 195-298. Dans cet article il est montré que pour un test d'hypothèse composite "à un côté" relatif à une famille de distributions exponentielles, une décision minimax est de la forme : Accepter l'hypothèse si $x < x_0$ et accepter l'alternative si $x > x_0$. M. Sobel et Chernoff "An essentially complete class of decision functions for certain standard sequential problems" Annals. Math. Stat. - Vol 24 (1953) pp. 319-337 ont obtenus quelques-uns de ces résultats pour la même classe de distributions lorsque la suite des décisions est finie.

Dans la première partie de l'article, S. KARLIN et H. RUBIN insistent sur les conditions à imposer aux fonctions pertes, les densités sont limitées et étudiées du point de vue de leurs propriétés. Dans la deuxième partie, les auteurs établissent

quelques lemmes fondamentaux, notamment ils mettent en évidence les changements de signes qui restreignent les propriétés des densités à rapport de vraisemblance monotone. Lorsque le nombre des décisions possibles est fini, la suite de toutes les stratégies monotones dans la classe des procédés statistiques est complète (partie 3). La quatrième partie est la détermination de la forme des stratégies de Bayes pour le statisticien tandis que le problème de l'admissibilité fait l'objet de la cinquième partie. La sixième partie réduit cette étude des stratégies pour la nature aux cas de deux actions. La partie sept est la recherche d'une classe complète lorsque le nombre des actions est fini. La partie huit est l'analyse de la nature des stratégies de Bayes pour un nombre fini d'actions. La dernière aborde une rapide discussion sur la théorie de l'invariance et les conditions de monotonie étudiées dans l'article.

ON THE POWER OF CERTAIN TESTS FOR INDEPENDANTE IN BIVARIATE POPULATIONS by M.S. KONIJN -

Soit F_{λ^o} la distribution à deux dimensions de deux variables indépendantes Y_{λ^o} et Z_{λ^o}. L'article est une étude de la distribution F_λ (à deux dimensions) de deux variables Y_λ et Z_λ transformées linéaires des deux premières. Soit J_0 le test de corrélation de rang de Spearman, J_1 le test de corrélation de signes, J_2 le test sans distorsion de corrélation de "graduation" J_3 le test de corrélation "médiale" et R le test de corrélation (paramétrique) ordinaire. L'auteur étudie quelques propriétés des puissances de ces tests pour des alternatives liées à l'hypothèse d'indépendance et pour de grands échantillons. Dans ces conditions l'efficacité de J_3 est fortement liée aux propriétés de localisation des densités de Y_{λ_0} et Z_{λ^0} ; l'efficacité de J_1 est souvent voisine de l'unité.

THE EFFICIENCY OF SOME NON PARAMETRIC COMPETITORS OF THE TEST - t - by J.L. HODGES

Soient des échantillons de deux distributions continues $F(x)$ et $F(x-\theta)$. On teste l'hypothèse $\theta = 0$ à l'aide du test de Wilcoxon. Les auteurs montrent dans le paragraphe 1, que l'efficacité asymptotique de ce test au sens de Pitman, relativement à celle du t, est toujours supérieure ou égale à 0,864. Ce résultat est encore valable pour le test de Kruskal - Wallis relativement au test F. Dans la deuxième partie, les auteurs comparent un certain nombre de notions d'efficacité asymptotique; alors surgissent certaines difficultés du fait que la puissance n'est pas nécessairement une fonction convexe de la taille de l'échantillon. En particulier à l'efficacité asymptotique de Pitman, les auteurs opposent une nouvelle notion basée sur la vitesse avec laquelle la puissance pour une alternative donnée tend vers 1. En particulier pour le test de signe relatif au test t et des populations normales, il est possible d'obtenir lorsque $n \longrightarrow \infty$ la limite de la suite des fonctions puissances.

ON REGULAR BEST ASYMPTOTICALLY NORMAL ESTIMATES - by CHIN LONG CHIANG -

Cette étude concerne l'estimation de certains paramètres trouvés lors de l'utilisation des processus stochastiques de croissance. Les distributions rencontrées dans ce domaine conduisent à des calculs différents et des formules complexes si l'on utilise les méthodes classiques d'estimation. Le but de cet article est de discuter une sous-classe des estimations envisagées par E. BARANKIN et J. GURLAND (1) et de donner une méthode simple pour construire de telles estimations. Ces estimations reposent sur un nombre de vecteurs aléatoires dont les distributions ne sont pas spécifiées. Il est démontré que, moyennant certaines conditions de régularité, les estimations régulières et convergentes ainsi obtenues sont asymptotiquement normalement distribuées lorsque le nombre des valeurs aléatoires tend vers l'infini.

(1) "An asymptotically Normal efficient estimators"
University of California publications in statistics - Vol. 1 pp.

L'auteur démontre la condition nécessaire et suffisante pour qu'un estimateur régulier et convergent ait une matrice des covariances asymptotiques "minimale". Il prouve également que si une fonction f satisfait certaines conditions, pour que $f(\hat{\theta})$ soit un estimateur RBAN de $f(\theta)$ à $f(\theta°)$, il est nécessaire et suffisant que $\hat{\theta}$ soit lui-même un estimateur RBAN de θ à $\theta°$. (R.B.A.N. = regular bestasymptotically normal).

THE LARGE SIMPLE POWER OF RANK IN THE TWO SAMPLE PROBLEM by M. DWASS -

Cet article étudie la puissance (pour de grands échantillons) de certains tests qui utilise la méthode des rangs contre des alternatives à un paramètre pour la comparaison de deux échantillons.

Les m premières, parmi N variables indépendantes, sont supposées identiquement distribuées avec comme densité $f_1(x, \theta)$; les N - m autres ont pour densité $f_2(x, \theta)$. Lorsque $\theta = 0$, les deux densités sont les mêmes. Si a_{N_1}, a_{N_2} ... a_{NN} est une suite de constantes, b_{N_1}, b_{N_2} ... b_{NN} une autre suite; si R_1, R_2 ... R_N sont les rangs des n variables, un statistique du type $\sum_{i=1}^{N} a_{N_i} b_{NR_i}$ est appelé un statistique L.

La partie I de cet article caractérise la statistique de rang localement la meilleure pour tester $H_0 = \theta = 0$ contre $H_1 = \theta > 0$ et voisin de 0. Moyennant certaines conditions de régularité, il est possible de déterminer la puissance des statistiques L et en particulier celle du statistique L localement le meilleur.

Si aucune condition n'est imposée aux b_{N_1} .. b_{NN}, il est habituellement difficile de déterminer si les conditions de régularité sont remplies; la partie II considère donc une classe spéciale de statistiques L, les statistiques L_h. Pour ceux-ci, les conditions de régularité sont plus faciles à vérifier. Le meilleur statistique L peut alors être approché à l'aide des statistiques L_h.

TABLES OF EXPECTED VALUES OF ORDER STATISTICS AND PRODUITS OF ORDER STATISTICS FOR SAMPLES OF SIZE TWENTY AND LESS FROM THE NORMAL DISTRIBUTION by D. TEICHROEW

Des tables des moyennes, variances et covariances des statistiques d'ordre provenant d'échantillons de 10 ou moins ont été données avec 5 décimales par GODWIN "On the estimation of dispersion by linear systematic statistics" Biometrica Vol. 36 (1949) - pp. 92-100. Dans cet article sont tabulées les valeurs moyennes, les écarts-type des statistiques d'ordre et produit de deux statistiques d'ordre pour des échantillons de 20 et moins avec 10 chiffres significatifs. De plus des fonctions nécessaires au calcul de ces mêmes valeurs moyennes à partir de grands échantillons sont fournies avec 25 chiffres significatifs.

ESTIMATION OF LOCATION AND SCALE PARAMETERS BY ORDER STATISTICS FROM SINGLY AND DOUBLY CENSORD SAMPLES - PART. I - THE NORMAL DISTRIBUTION UP TO SAMPLES OF SIZE by A.E.SARHAN et B.C.GREENBERG.

Des échantillons censurés du type II font l'objet de cet article; la taille de l'échantillon est connue mais les observations de plusieurs valeurs extrêmes manquent. Les échantillons simplement censurés sont ceux pour lesquels font défaut seulement les r_1 plus petites observations ou les r_2 plus grandes; ils sont doublement censurés lorsque ces r_1 et r_2 valeurs manquent. La partie I fournit des tables permettant l'estimation de la moyenne et de l'écart-type d'une distribution normale à partir d'échantillons simplement et doublement censurés. (taille des échantillons $n \leq 10$).

Les estimations des paramètres utilisés pour construire les estimateurs systématiques linéaires sont obtenues en rangeant par ordre de grandeur croissante les éléments connus de l'échantillon et en appliquant la méthode des moindres carrés pour obtenir la meilleure combinaison linéaire de ces paramètres. Les coefficients ainsi calculés font les estimations sans distorsion et à variance minimum. La mé-

thode du calcul utilisé est celle de Gupta avec quelques modifications (A.K. GUPTA "Estimation of the mean and standand deviation of a normal population from a censored sample" Biometrica Vol. 39 (1954).

A WAITING LINE PROCESS OF MARKOFF TYPE -
by A.B. CLARKE -

Les processus des queues du type de Markoff font l'objet de cet article. Le traffic des "arrivées" est du type de Pearson. Les paramètres sont fonctions du temps. La recherche d'une solution exacte pour l'obtention de la distribution de la longueur de la file d'attente fonction du temps se réduit à la résolution d'une équation de Voltera. Lorsque le rapport des paramètres relatifs aux entrées et sorties est constant, cette équation se résout explicitement et on peut obtenir la distribution. Apartir de cette solution, on peut étudier le processus pour des grandes valeurs de t, particulièrement dans le cas instable lorsque l'intensité est supérieure à 1.

ON THE NORMAL APPROXIMATION TO THE HYPERGEOMETRIC DISTRIBUTION
by W.L. NICHOLSON

Dans cet article, l'auteur fournit une nouvelle approximation de la somme des termes d'une distribution hypergéométrique; c'est une généralisation de l'approximation normale donnée par Feller pour la distribution binomiale. Lorsque la distribution est asymétrique cette approximation est une amélioration par rapport aux résultats classiques. L'auteur en déduit des limites inférieures et supérieures pour la distribution hypergéométrique ce qui permet d'évaluer l'importance de l'erreur.

CHARTS OF THE POWER OF THE F TEST BY M. FOX

Les coordonnées sont f_1 et f_2 nombres de degrés de liberté des numérateurs et dénominateurs du statistique F; les cartes sont construites pour $\beta = 0.5, 0.7, 0.8, 0.9$ avec $\alpha = 0.001$ et $\alpha = 0.05$. Les courbes sont valables pour des valeurs de φ = constante ou $\varphi = \sqrt{\delta_b^2/(f_1+1)\sigma^2}$. et δ_b^2 est le numérateur de F.

Ces cartes sont destinées à répondre à la question suivante : Quelle combinaison de f_1 et f_2 doit-on utiliser de manière à ce que la puissance ait un niveau fixé β ?

ESTIMATING THE PARAMETERS OF A TRUNCATED GAMMA DISTRIBUTION
by D.G. CHAPMAN

Dans cet article, l'auteur publie une table permettant de simplifier l'estimation des paramètres d'une fonction gamma incomplète ou distribution du type III. La nouvelle méthode ainsi suggérée peut être utilisée pour un certain nombre d'autres distributions tronquées, que la coupure porte sur le centre ou les extrémités de ces distributions.

APPROXIMATE UPPER PERCENTAGE POINTS FOR EXTREM VALUES IN MULTINOMIAL SAMPLING by R.M. KOZELKA

Etant donnée une distribution k - multinomiale avec des probabilités égales pour chaque catégorie, on cherche la probabilité de la fréquence la plus grande dans chaque catégorie. L'article alors donne une approximation asymptotique simple des centiles supérieurs de cette distribution. Une table fournit les points 0,95 et 0,99 calculés à partir de cette approximation pour k = 1 (1) 25 est jointe une table permettant de comparer ces valeurs à celles réelles, lorsque k = 2, 4, 5 et n = 3(1) 12

AN EXTENSION OF THE KOLMOGOROFF DISTRIBUTION
by J. BLAKMAN

Soit x_1, x_2, ..., x_n, x'_1 ... x'_n des variables aléatoires indépendantes avec une distribution commune $F(x)$ continue. Soit $F_n(x)$ la distribution empirique construite à l'aide de x_1, x_2, ..., x_n, $G_n(s)$ celle construite à partir de x'_1, ... x_n. L'auteur calcule les valeurs exactes de $Pr[-y < F_n(s) - G_n(s) < x$ pour tout $s]$ et $Pr[-y < F(s) - F_n(s) < x$ pour tout $s]$ ainsi que les deux premiers termes des séries asymptotiques lorsque n est grand.

ON THE SIMULTANEOUS ANALYSIS OF VARIANCE TEST -
by K.V. RAMACHANDRAN

Dans l'analyse de la variance les différents tests de comparaison des moyennes ne sont pas indépendants. La théorie des tests simultanés de l'analyse de la variance et son utilisation pour certains problèmes ont été données par CHOSH "Simultaneous tests of linear hypotheses by the method of analysis of variance" (unpublished 1953). L'auteur n'envisage que certains cas particuliers et montre que la puissance du test à des propriétés de monotonie.

J. & R. SENNAC, Imprimeurs
54, Fg Montmartre, PARIS (9e)